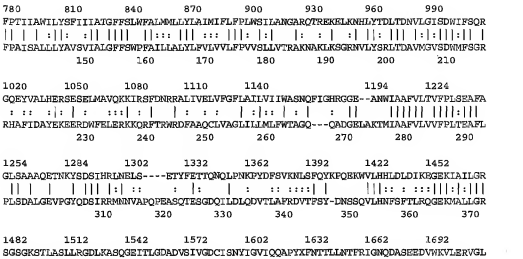
```
Query: 368 HLDLDIKEGEKIAILGRSGSGKSTLASLLRGDLKASQGRITLGDADVSIVGDCISNYIGV 427
           +     +++GEK+A+LGRSGSGKST   +L+ G LK    G +TL    + +++ D I++ + V
Sbjct: 355 NFSFTLRQGEKMALLGRSGSGKSTSLALIEGALKPDSGSVTLNGVETALLKDQIADAVAV 414

Query: 428 IQQAPVLFNTTLLNNIRIGNQDASEDVWKVLERVGLKGMVTDLSDGLYTNVDEAGLRPSS 487
           + Q P+LF+T++LNNIR+GN +AS+EDV +  +V L + +  L DG +T V E G+RPS
Sbjct: 415 LNQKPHLFDTSILNNIRLGNGEASDEDVRRAAKQVKLHDYIESLPDGYNTSVQETGIRPSS 474

Query: 488 GGERHRIALARILLKDVPIVILDEPTVGLDPITEQALLRVFMKKLEGKTLVWITHHLKGI 547
           GGER RIALARILL+D PI+ILDEPTVGLDPITE+ L+     + L+GKT++WITHHL G+
Sbjct: 475 GGERQRIALARILLQDTPIIILDEPTVGLDPITRERLMETVFEVLKGKTILWITHHLAGV 534

Query: 548 EHADRILFIENGQLELEGSPQELSQSSQRYRQL 580
           E AD+I+F+ENG+ E+EG+ +EL   +++RYR+L
Sbjct: 535 EAADKIVFLENGKTEMRGTHEELLAANERYRRL 567
```

A related GBS gene <SEQ ID 8861> and protein <SEQ ID 8862> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Cremd: 8
McG: Discrim Score:    -15.90
GvH: Signal Score (-7.5): 1.97
     Possible site: 49
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 7 value: -12.84 threshold:  0.0
   INTEGRAL   Likelihood =-12.84   Transmembrane  260 - 276 ( 258 - 284)
   INTEGRAL   Likelihood = -9.34   Transmembrane  172 - 188 ( 147 - 199)
   INTEGRAL   Likelihood = -6.53   Transmembrane  150 - 166 ( 147 - 171)
   INTEGRAL   Likelihood = -6.05   Transmembrane   31 -  47 (  29 -  52)
   INTEGRAL   Likelihood = -3.35   Transmembrane   68 -  84 (  67 -  84)
   INTEGRAL   Likelihood = -1.17   Transmembrane  293 - 309 ( 292 - 310)
   INTEGRAL   Likelihood = -0.69   Transmembrane  494 - 510 ( 493 - 510)
   PERIPHERAL Likelihood =  3.29     412
 modified ALOM score:   3.07

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.6137(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00997(346 - 2052 of 2364)
EGAD|98910|BS3866(1 - 571 of 575) transport ATP-binding protein cydd {Bacillus subtilis}
OMNI|NT01BS4517 ABC transporter CydC, putative SP|P94367|CYDD_BACSU TRANSPORT ATP-BINDING
PROTEIN CYDD. GP|1783253|dbj|BAA11730.1||D83026 homologous to many ATP-binding transport
proteins; hypothetical {Bacillus subtilis} GP|2636408|emb|CAB15899.1||Z99123 ABC membrane
transporter (ATP-binding protein) {Bacillus subtilis} PIR|D69611|D69611 ABC transporter
required for expression of cytochrome bd (ATP-) cydD - Bacillus subtilis
%Match = 31.9
%Identity = 45.2  %Similarity = 69.1
Matches = 257  Mismatches = 172  Conservative Sub.s = 136

          300       330       360       390       420       450       480       510
LKKDISIN*SMLWEEMMPKIPLPKRLKTDQWIKPFPKQYKVSLVIALPLGFWTPPSASALMFNSGYLISKSASLPSNILL
                                :| ::|| |: ||    :|: :||| :| |||: ||| ||:|:|||:|: | ||||
                               MKKREWILPYIKQNARLPVLVIPLGAVTIPSAAFLMFTSGFLISKAATRPENILL
                                         10        20        30        40        50

          540       570       600       630       660       690       720       750
VYVPIVLVRAFGIGKPVFRYIERLTSHNWVLRMTSQLRLKDLYHSLESNAIFMKKDPRLGDVMGLLAEDINYLQNLYLRTI
:|||||  | ||| | | ||:|||  |: ::|:: | :|::||: ||  |: ::  || ||::|:|:||| :|: :|:||
IYVPIVAVRTPGIARSVSRYVERLVGHHITLKIVSLMRVRLYHMLEPGALMLRSRPFTGDMLGILSEDIEHLQDAFLKTI
          70        80        90       100       110       120       130
```

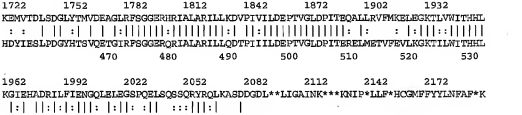


There is also homology to SEQ ID 478.

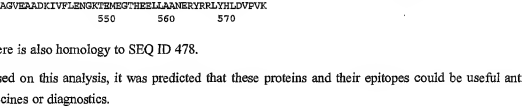
Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1616

A DNA sequence (GBSx1711) was identified in *S.agalactiae* <SEQ ID 4987> which encodes the amino acid sequence <SEQ ID 4988>. This protein is predicted to be spore germination protein C3 (ispB).



The protein has homology with the following sequences in the GENPEPT database.



```
Query: 68  QQLKKIAASLEILHVATLIHDDVIDDSPLRRGNMTIQSKFGKDIAVYTGDLLFTVFFDLI 127
            ++K +A +LE++H+A+L+HDDVIDD+ LRRG  TI++K+    IA+YTGD +      +++
Sbjct: 92  NKIKYVAVTLEMIHMASLVHDDVIDDAKLRRGKPTIKAKWDNRIAMYTGDYMLAGSLEMM 151

Query: 128 LESMADTPFMRINAKSMRKILMGELDQMHLRYNQQQGIHHYLRAISGKTAELFKLASKSG 187
             + +    RI ++++ ++ +GE++QP  +YN +Q +  YLR I  KTA L  ++ + G
Sbjct: 152 TR-INEPKAHRILSQTIVEVCLGEIEQIKDKYNMEQSLRTYLRRIRRKTALLIAVSCQLG 210

Query: 188 AYPGGARKEVVRLAGHIGFNIGMTFQILDDILDYTADKKTFNKPVLEDLAQGTYSLPLLL 247
            A   GA++++ +      G+ +GM++QI+DDILD+T+ ++    KPV  DL QG  +LP+L
Sbjct: 211 ATASGADEKIHKALYWFGYYVGMSYQIIDDILDFTSTEKELGKPVGGDLLQGNVTLPVLY 270

Query: 248 AIEENPDIFKPLLDKKTDMATEDMEKIAYLVVSHRGVDKARHLARKFTEKAISDINKLPQ 307
            A+ +NP +   +     ++   E +E I   +      ++ +  ++  + +KA   +N LP+
Sbjct: 271 AL-KNPALKQLKLINSETTQEQLEPIIEEIKKTDATEASMAVSEMYLQKAFQKLNTLPR 329

Query: 308 SSAKKQLLQLTNYLLKRK 325
            A+  L  +  Y+ KRK
Sbjct: 330 GRARSSLAAIAKYIGKRK 347
```

There is also homology to SEQ ID 284. An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 65/227 (28%), Positives = 98/227 (42%), Gaps = 9/227 (3%)

Query: 43  AGGKQLRPAFFYLFSQLGNKENQDTQQLKKIAASLEILHVATLIHDDV--IDDSPLRRGN 100
            +GGK++RP       +      Q+      +AA+LE++H  +LIHDD+  +D+    RRG
Sbjct: 36  SGGKRIRPLILLEMIEGPGVSLQNAHF--DLAAALEMIHTGSLIHDDLPAMDNDDYRRGR 93

Query: 101 MTIQSKFGKDIAVYTGDLLFTVFFDLILESM--ADTPFMRINAKSMRKILMGELDQMHLR 158
            +T   +FG+  A+  GD LF   F LI ++    +      I   S+    G +    L
Sbjct: 94  LINHKQPGEATAILAGDSLFDLPFGLILAQAELNSEVKVALIQELSLASGTPGMVGGQMLD 153

Query: 159 Y---NQQQGIHHYLRAISGKTAELFKLASKEGAYFGGAKEVVRLAGHIGFNIGMTFQILL 215
               NQ   +      KT +L     K A       V +     G  IG   FQI
Sbjct: 154 MKGENQALSLPQLSLIHLNKTGKLLAPPFKAAALITEQAMTVRQQLEQAGMLIGHAPQIR 213

Query: 216 DDILDYTADKKTFNKPVLEDLAQGIYSLPLLLAIEENPDIFKPILDK 262
            DDILD TA  +   K   +DL     + P LL +E +  +    LD+
Sbjct: 214 DDILDVTASFEDLGKTPKKDLFAEKATYPSLLGLEASYQLLTESLDQ 260
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1617

A DNA sequence (GBSx1712) was identified in *S.agalactiae* <SEQ ID 4989> which encodes the amino acid sequence <SEQ ID 4990>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3995(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25232 GB:M58315 dipeptidyl peptidase IV [Lactococcus lactis]
 Identities = 385/767 (50%), Positives = 504/767 (65%), Gaps = 21/767 (2%)

Query: 1   MRYNQFSYIPTKDNEAFEELKGLGPFLNKKNSDKANLEAFLRHSFLNQTDTDYALSLLIV 60
            MR+N FS +    +E   EL  LGF  +    +K  L+ FL  S  + TD       L
Sbjct: 1   MRFNHFSIVDKNFDRQLAELDQLGPFWSVPWDEKKILKDFLIQSPSDMTD-------LQA 53
```

```
Query: 61  DAKTDALTFFKSNSDLTLENLQWIYLQLLGFIPFVDFKDPKAF-------LQDINFPVSY 113
           A+ D + F KS+ +L  E    I LQLL F+P  DF+  KAF       L  I   ++
Sbjct: 54  TAELDVIEFLKSSIELDWEIPWNIALQLLDFVPNFDPEIGKAFEYAKNSNLPQIRAEMTT 113

Query: 114 DNIPQSLHHLLACRGKSGNTLIDQLVADGLLHADNEYHFPNGKSLATFNTNQLIREVVYV 173
           +NI  + ++LL  R K+G  L++  V++GLL  DNEYHFFN KSLATF+++ L REV++V
Sbjct: 114 ENIISAFYYLLCTRRKNGMILVEHHVVSEGLLPLDNEYHFFNDKSLATFDSSLLEKEVLNV 173

Query: 174 ETSLDTMSSGEHDLVKVNIIRPTTEHTIPTWMTASPVHQGINDPAADQKTYQMEGALAVK 233
           E+ +D+  GE+DL+K+ IIRP +  +P +MTASPVH GIND A D  + M  L K
Sbjct: 174 ESFVDSEQRGKNDLIKIQIIRPKSTEKLPVVMTASPVHIGINDKANDLALHDMNVELEEK 233

Query: 234 QPKIIQVDTKPFKEEVKHPSKLPI-SPATSSFTHIDSYSLNDYFLSRGFANIYVSGVSTA 292
              I V+ K  ++      +LPI   A    FTH  +YSLNDYFL+RGFA+IYV+GVGT
Sbjct: 234 TSHEIHVEQKLFQKLSAKAKELPIVDKADVRFTHGWTYSLNDYFLTRGFASIYVAGVGTR 293

Query: 293 GSTGFMTSGDYQQIQSFKAVIDWLNGKVTAFTSHKRDKQVKANWSNGLVATTGKSYLGTM 352
           S GF TSGDYQQI S  AVIDWLNG+  A+TS K+  ++KA+W+NG VA TGKSYLGTM
Sbjct: 294 SSDGFQTSGDYQQIYSMTAVIDWLNGRARAYTSRKKTHEIKASWANGKVAMTGKSYLGTM 353

Query: 353 STGLATTGVESLKVIIAEAAISTWYDYYRENGLVCSPGGYPGEDLDVLTELTYSRNLLAG 412
           + G ATTGVESL+VI+AEA IS+WY+YYRENGLV SPGG+PGEDLDVL  LTYSRNL
Sbjct: 354 AYGAATTGVESLEVIIAEAGISSWYNYYRENGLVRSPGGFPGEDLDVLAALTYSRNLDGA 413

Query: 413 DYIKNNDCYQALLNEQSKAIDRQSGDYNQYWHDRNYLTVNNVKSRVVYTHGLQDWNVKP 472
           D++K N  Y+  L E + A+DR+SGDYNQ+WHDRNYL + + VK+ V+  HGLQDWNV P
Sbjct: 414 DFLKGNAESYEKRLAEMTAALDRKSGDYNQFWHDRNYLINTDEVKADVLIVHGLQDWNVTP 473

Query: 473 RHVYKVFNALPQTIKKHLFLHQGQHVYMHNWQSIDFRESMNALLSQELLGIDNHFQLEEV 532
              Y  + ALP+   KH FLH+G H+YM++WQSIDF E++NA    +LL  D +  L  V
Sbjct: 474 EQAYNFWKALPEGEAKHAFLHRGAHIYMNSWQSIDFSETINAYFVAKLLDRDLENLNLPPV 533

Query: 533 IWQDNTTEQTWQVLDAFGGNHQEQIGLGD---SKKLIDNHYDKEAPDTYCKDFNVFKNDL 589
           I Q+N+ +Q W +++ FG N Q ++ LG    S   DNHYD E F  Y KDFNVFK DL
Sbjct: 534 ILQENSKDQVWIWMNDFGANTQIKLPLGKTAVSPAQFDNHYDDETFKKYSKDFNVFKDDL 593

Query: 590 FKGNNKTNQITINLPLKKNYLLNGQCKLHLRVKTSDKKAILSAQILDYGPKKRFKDTPTI 649
           F+  NK N+  I+L L      +NG  +L LR+K +D K  LSAQILD+G KKR +D   +
Sbjct: 594 FE--NKANEAVIDLELPSMLTINGPVELELRLKLNDTKGFLSAQILDFGQKKRLEDKARV 651

Query: 650 KFLNSLDNGKNFAREALRELPFTKDHYRVISKGVLNQNRTDLLTIEAISPEQWFDIEFS 709
           K   LD G+NF + L ELP  +  Y++I+KG  NLQN+ +LIT+  ++ ++WF I+F
Sbjct: 652 KDFKVLDRGRNFMLDDLVELPLVESPYQLITKGFTNLQNQ-NLLTVSDLKADEWFTIKFE 710

Query: 710 LQPSIYQLSKGDNLRIILYTTDFEHTIRDNASYSITVDLSQSYLTIP 756
           LQP+IY L K D LR+ILY+TDFEHT+RDN   +  +DLSQS L IP
Sbjct: 711 LQPTIYHLEKADKLRVILYSTDFEHTVRDNRKVTYEIDLSQSKLIIP 757
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4991> which encodes the amino acid sequence <SEQ ID 4992>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2553(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 481/758 (63%), Positives = 587/758 (76%), Gaps = 4/758 (0%)

Query: 1   MRYNQFSYIPTKPNEAPEELKGLGFPLNKKNSDKANLEAFLRHSPLNQTDTDYALSLLIV 60
           MRYNQFSYIPT    A EELK LGF L+ + + KA+LE+FLR  F +  D+DY LS LI
Sbjct: 1   MRYNQFSYIPTSLERAAEELKSLGFDLDLQKTAKASLESFLRKLFFHYPDSDYFLSHLIA 60
```

```
Query: 61  DAKTDALTPFKSNSDLTLENLQWIYLQLLGFIPFVDFKDPKAFLQDINFPVSYDN--IFQ 118
           DAL+FF+S  +L+ E    + LQ+LGFIP VDF +  AFL  + FP+ +D   I +
Sbjct: 61  KNDMDALSFFQSEQELSKEVFDLLALQVLGFIPGVDPTEADAFLDKLAFPIHFDETEIIK 120

Query: 119 SLHHLLACRGKSGNTLIDQLVADGLLHADNHYHFPNGKSLATFNTQLIREVVYVETSLD 178
            +HHLLA R KSG TLID LV+ G+L  DN YHFPNGKSLATF+T+QLIREVVYVE  LD
Sbjct: 121 HIHHLLATRCKSGMTLIDDLVSQGMLTMDNDYHFPNGKSLATFDTSQLIREVVYVEAPLD 180

Query: 179 TMSSGEHDLVKVNIIRPTTEHTIPTWMTASPYHQGINDPAADQKTYQMEGALAVKQPKHI 238
           T   G+ DL+KVNIIRP ++  +PT+MT SPYHQGIN+ A D+K Y+ME  L VK+ + I
Sbjct: 181 TDQDSQLDLIKVNIIRPQSQKPLPTLMTPSPYHQGINEVANDKKLYRMEKKLVVKKRQI 240

Query: 239 QVDTKPFKEEVKHPSKLPISPATESFTHIDSYSLNDYFLSRGFANIYVSGVGTAGSTGFM 298
            V+ + F      P KLPI    ESF++I+SYSLNDYFL+RGFANIYVSGVGTAGSTGFM
Sbjct: 241 TVSDRDFIPLETQPCKLPIGQNLESFSYINSYSLNDYFLARGFANIYVSGVGTAGSTGFM 300

Query: 299 TSGDYQQIQSFKAVIDWLNGKVTAPTSHKRDKQVKANWSNGLVATTGKSYLGTWSTGLAT 358
            TSG+Y QI+SFKAVIDWLNG+ TA+TSH +  QV+A+W+NGLV TTGKSYLGTWSTGLAT
Sbjct: 301 TSGNYAQIESFKAVIDWLNGRATAYTSHSTHQVRADWANGLVCTTGKSYLGTWSTGLAT 360

Query: 359 TGVEGLKVIIAEAAISTWYDYYRENGLVCSPGGYPGEDLDVLTELTYSRNLLAGDYIKNN 418
            TGV+GL +IIAE+AISWY+YYRENGLVCSPGGYPGEDLDVLTELTYSRNLLAGDY+++N
Sbjct: 361 TGVDGLAMIIAESAISSWYNYYRENGLVCSPGGYPGEDLDVLTELTYSRNLLAGDYLRHN 420

Query: 419 DCYQALLNEQSKAIDRQSGDYNQYWHDRNYLTHVNNVKSRVVYTHGLQDMNVKPRHVYKV 478
            D YQ LLN+QS+A+DRQSGDYNQ+WHDRNYL + + +K  VVYTHGLQDMNVKPR VY++
Sbjct: 421 DRYQELLNQQSQALDRQSGDYNQFWHDRNYLKNAHQIKCDVVYTHGLQDMNVKPRQVYEI 480

Query: 479 FNALPQTIKKHLFLHQGQHVYMHNWQSIDFRESMNALLSQELLGIDNHFQLEEVIWQDNT 538
            FNALP TI KHLFLHQG+HVYMHNWQSIDFRESMNALL Q+LLG+ N F L E+IWQDNT
Sbjct: 481 FNALPSTINKHLFLHQGEHVYMHNWQSIDFRESMNALLCQKLLGLANDFSLPEMIWQDNT 540

Query: 539 TEQTWQVLDAFGGNHQEQIGLGDSKKLIDNHYDEEAFDTYCKDFNVFKNDLFKGNNKTNQ 598
            Q WQ    FG +  +++ LG    LIDNHY ++ F  Y KDF  FK  LFKG  K NQ
Sbjct: 541 CPQNWQERKVFGTSTIKELDLGQELLLIDNHYGEDEPKAYGKDFRAFKAALFKG--KANQ 598

Query: 599 ITTNLPLKKNYLLNGQCKLHLRVKTSDKKAILSAQILDYGPKKRFKDTPTIKFLNSLDNG 658
             I++ L+++  +NG+  L L+VK+S+ K +LSAQILDYG KKR  D P     +S+DNG
Sbjct: 599 ALIDILLEEDLPINGEIVLQLKVKSSENKGLLSAQILDYGKKKRLGDLPIALTQSSIDNG 658

Query: 659 KNFAREALRELPFTKDHYRVISKGVLNLQNRTDLLTIEAIEPEQWFDIEFSLQPSIYQLS 718
            +NF+RE L+ELPF +D YRVISKG +NLQNR +L  +IE I   +W +   LQP+IY L
Sbjct: 659 QNFSREPLKELPFREDSYRVISKGFMNLQNRNNLSSIETIPNNKWMTVRLPLQPTIYHLE 718

Query: 719 KGDNLRIILYTTDFEHTIRDNASYSITVDLSQSYLTIP 756
            KGD LR+ILYTTDFEHT+RDN++Y++T+DLSQS L +P
Sbjct: 719 KGDTLRVILYTTDFEHTVRDNSNYALTIDLSQSQLIVP 756
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1618

A DNA sequence (GBSx1713) was identified in *S.agalactiae* <SEQ ID 4993> which encodes the amino acid sequence <SEQ ID 4994>. This protein is predicted to be PrfA. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3976(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10125> which encodes amino acid sequence <SEQ ID 10126> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA65740 GB:X97014 PrfA [Listeria seeligeri]
 Identities = 54/181 (29%), Positives = 95/181 (51%), Gaps = 1/181 (0%)

Query: 38  DYTYILKDGIVKQSVLSKYGTEFNLRYVTGLKITSILNTDYSQHMGRPYNVRIESETAHF 97
            +Y   L +G+ K + +S+ G   NL+Y  G  I      D  + +G  YN+ + SE A
Sbjct: 36  EYCIFLHEGVAKLTSISRSGDILNLQYYKGAFIIMTGFIDTEKSLGY-YNLEVVSEQAAA 94

Query: 98  YKVRRSTFLKDINNDIELQGYVKDFYHNRLEKSMKHMQCMLTNGRIGAISTQLYDLSKMF 157
            Y ++ S   + ++ D++   Y+ D    ++  S+ K      +NG++G+I  Q   L+ ++
Sbjct: 95  YIIKISDLKELVSKDMQLFYIIDTLQKQVSYSLAKFNDFSSNGKVGSICGQFLILAVYY 154

Query: 158 GEERDNGDIYINFVITNEELGKFCGISTGSSVSRILKQLKDDHIIRIEKQHIIITNVEKLK 218
            GEE  NG       +T +ELG   GI+  S+VSRI+ +LK +++I  +  + I N+  LK
Sbjct: 155 GEETPNGIKITLEKLTMQELGCSSGIAHSSAVSRIISKLKQENVIEYKDSYFYIKNIAYLK 215
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4995> which encodes the amino acid sequence <SEQ ID 4996>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4088(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 186/223 (83%), Positives = 203/223 (90%)

Query: 1   MEEVMNHQILQNYINSHNLPIIEKDYHKYLTFESLEEDYTYILKDGIVKQSVLSKYGTEF 60
           +E+ +NH ILQ YI++HN PIIEK YHKYLTFESLEED+TYILKDGIVKQSVLSKYG EF
Sbjct: 17  LEKSVNHHILQRYIDHNNPPIIEKSYHKYLTFESLEEDFTYILKDGIVKQSVLSKYGMEF 76

Query: 61  NLRYVTGLEITSILNTDYSQHMGEPYNVRIESETAHFYVKVRRSTFLKDINNDIELQGYVK 120
           NLRYVTGLEITS+LNT YS+ MGEPYNVRIESE A FYKVRRS FLKDIN DIELQGYVK
Sbjct: 77  NLRYVTGLEITSVLNTGYSKDMGEPYNVRIESEKASFYKVRRSAFLKDINRDIELQGYVK 136

Query: 121 DFYHNRLEKSMKHMQCMLTNGRIGAISTQLYDLSKMFGEERDNGDIYINFVITNEELGKF 180
           DFYHNRL+KSMKHMQCMLTNGRIGAISTQ+YDL  +FGEE  NG I INFVITNEELGKF
Sbjct: 137 DFYHNRLQKSMKHMQCMLTNGRIGAISTQIYDLMTLFGEELPNQQILINFVITNEELGKF 196

Query: 181 CGISTGSSVSRILKQLKDDHIIRIEKQHIIITNVEKLKDHIVF 223
           CGIST SSVSRILKQLK+ +IRI+KQHIIITN++KLKD+IVF
Sbjct: 197 CGISTASSVSRILKQLKKNNIRIDKQHIIITNLDKLKDNIVF 239
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1619

A DNA sequence (GBSx1714) was identified in *S.agalactiae* <SEQ ID 4997> which encodes the amino acid sequence <SEQ ID 4998>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.33   Transmembrane  167 - 183 ( 159 - 193)
   INTEGRAL    Likelihood = -7.96   Transmembrane   18 -  34 (  10 -  37)
   INTEGRAL    Likelihood = -7.75   Transmembrane  373 - 389 ( 369 - 392)
```

```
INTEGRAL    Likelihood = -5.68   Transmembrane  214 - 230 ( 212 - 234)
INTEGRAL    Likelihood = -4.78   Transmembrane  243 - 259 ( 241 - 262)
INTEGRAL    Likelihood = -2.71   Transmembrane   48 -  64 (  47 -  65)
INTEGRAL    Likelihood = -2.60   Transmembrane  283 - 299 ( 283 - 300)

----- Final Results -----
              bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15662 GB:Z99122 similar to antibiotic resistance protein
        [Bacillus subtilis]
 Identities = 106/401 (26%), Positives = 199/401 (49%), Gaps = 21/401 (5%)

Query: 3   DKLFVNKHFIGITILNPIVYMVYLFTVIIAFIATKELGVSTSQAGLATGIYIVGTLIARL 62
            D ++ K FI + ++N  V++ +Y F  ++      +ELG + SQ GL   ++++  +I R
Sbjct: 5   DAIWTKDFIMVLLVNLFVFVFFTTFLTVLPIYTLQELGTESQGGLLISLFLLSAIITRP 64

Query: 63  IPGKQLEVLGRXLVLRGGAIFVLLTTLAYFVMPSIGVMVLVRVLNGRGYGVVSTXTNTIV 122
                +E  G+K +       + L++  Y  + +  ++  +RF  G  + +++T T  I
Sbjct: 65  FSGAIVERFGKKWMAIVSMALFALSSFLVMPIHNFSLLLGLRFFQGIWFSILTTVTGAIA 124

Query: 123 TAYIPADKRGEGINFYGLSTSLAAAIGPFVGTFMLLNLHINFKMVIVLCSILIAIVVLGA 182
                IPA +RGEG+ ++ +S +LA AIGPF+G  ++     ++F +      ++ +    +L +
Sbjct: 125 ADIIPAKRRGEGLGYFAMSMNLAMAIGPFLGLNLMRV--VSFFVFPTAFALFMVAGLLVS 182

Query: 183 FVFPVKNITLNPEQLAKSKSWTIDSF-----IEKKAIFITTIAFLMGISYASVLGFQKLY 237
            F+  V          +K     T+  F       EK A+ I  +   +  Y++V  +  ++
Sbjct: 183 FLIKVPQ--------SKDSGTTVFRFAFSLMFEKGALKIATVGLFISFCYSTVTSYLSVF 234

Query: 238 TTEINLMTVGAYFFIVYALVITLTRPSMGRLMDAKGDKWVLYPSYLFLTLGLALLGSAMS 297
              ++L +  YFP+ +A+ + + RP  G+L D  G    V+YPS L  ++GL +L
Sbjct: 235 AKSVDLSDISGYFFVCFAVTMMIARPFTGKLFDKVGPGIVIYPSILIFSVGLCMLSFTHS 294

Query: 298 SVTYLLSGALIGFGYGTFMSCGQAASIKGVEEHRFNTAMSTYMIGLDLGLGAGPYILGLV 357
             +  LLSGA+IG GYG+ + C Q  +I+     HR   A +T+     D G+  G Y+ GL
Sbjct: 295 GLMLLLSGAVIGLGYGSIVPCMQTLAIQKSPAHRSGFATATFFTFFDSGIAVGSYVFGL- 353

Query: 358 KDGFLGAGVQSFRELFWIAAIIPVVCGILYFLKSSRQVETK 398
            F+ +    F  ++  A +  ++LY     + E +
Sbjct: 354 ---FVASA--GFSAIYLTAGLFVLIALLLYTWSQKKPAEAE 389
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4999> which encodes the amino acid sequence <SEQ ID 5000>. Analysis of this protein sequence reveals the following:

```
      Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.31   Transmembrane  202 - 218 ( 194 - 225)
   INTEGRAL    Likelihood = -7.80   Transmembrane   53 -  69 (  44 -  71)
   INTEGRAL    Likelihood = -7.17   Transmembrane  407 - 423 ( 404 - 426)
   INTEGRAL    Likelihood = -5.26   Transmembrane  249 - 265 ( 247 - 269)
   INTEGRAL    Likelihood = -3.77   Transmembrane  279 - 295 ( 276 - 297)
   INTEGRAL    Likelihood = -2.23   Transmembrane   11 -  27 (  10 -  27)
   INTEGRAL    Likelihood = -2.13   Transmembrane   83 -  99 (  82 -  99)
   INTEGRAL    Likelihood = -1.91   Transmembrane  312 - 328 ( 311 - 328)

----- Final Results -----
              bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15662 GB:Z99122 similar to antibiotic resistance protein
        [Bacillus subtilis]
```

```
Identities = 110/390 (28%), Positives = 194/390 (49%), Gaps = 11/390 (2%)

Query: 38  EKLFNKHFVAITVINFIVYMVYYLFTVIIAFVATRELGAQTSQAGLATGIYILGTLLARL 97
            + ++ K F+ + ++N  V++ +Y F  ++      +ELG   SQ GL    +++L  ++ R
Sbjct: 5   DAIWTKDFIMVLLVNLFVFVFFYTFLTVLPIYTLQELGGTESQGGLLISLFLLSAIITRP 64

Query: 98  IFGKQLEVFGRRLVLRGGAIFYLLTTLAYFYMPTISMMYLVRFLNGFGYGVVSTATNTIV 157
              G  +E FG++ +         + L++  Y  +   S++  +RF  G  + +++T T  I
Sbjct: 65  FSGAIVERFGKKRMAIVSMALFALSSFLYMPIHNFSLLLGLRFFQGIWFSILTTVTGAIA 124

Query: 158 TAYIPARKRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHIDFRMIIVLCSVLIGCVVVGA 217
                IPA++RGEG+ ++ +S +LA AIGPF+G  ++     + F +      ++ +    ++ +
Sbjct: 125 ADIIPAKRRGEGLGYFAMSMNLAMAIGPFLGLNLMRV--VSFPVFFTAFALFMVAGLLVS 182

Query: 218 FAFPVKNMSLNAEQLAKTKSWTVDSFIEKKALFITAIAFLMGIAYASVLGFQKLYTSEIH 277
            F   V      +   + +   +        EK AL I  +    +   Y++V  +  ++   +
Sbjct: 183 FLIKVPQSKDSGTTVFR---FAFSDMFEKGALKIATVGLFISFCYSTVTSYLSVFAKSVD 239

Query: 278 LTTVGAYFFVVYALIITITRPAMGRLMDAKGDKWVLYPSYLFLAMGLFLLGSVSSGGSYL 337
            L+ +  YFFV +A+ + I RP  G+L D  G   V+YPS L  ++GL +L    SG   L
Sbjct: 240 LSDISGYFFVCFAVTMMIARPFTGKLFDKVGPGIVIYPSILIFSVGLCMLSFTHSGLMLL 299

Query: 338 LSGALIGFGYGTFMSCGQAASIQGVDEHRFNTAMSTYMIGLDLGLGAGPYLLGLIKDLAL 397
            LSGA+IG GYG+ + C Q  +IQ     HR   A +T+     D G+  G Y+ GL
Sbjct: 300 LSGAVIGLGYGSIVPCMQTLAIQKSPAHRSGFATATFFTFFDSGIAVGSYVFGLF----- 354

Query: 398 GSGVASFRHLFWLAAVIPLICTLLYLLKTK 427
                A F  ++  A +  LI  LLY    K
Sbjct: 355 -VASAGFSAIYLTAGLFVLIALLLYTWSQK 383
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 328/396 (82%), Positives = 370/396 (92%), Gaps = 1/396 (0%)

Query: 1   MEDKLFNKHFIGITILNFIVYMVYYLFTVIIAFIATKELGVSTSQAGLATGIYIVGTLIA 60
            ME+KLFNKHF+ IT++NFIVYMVYYLFTVIIAF+AT+ELG  TSQAGLATGIYI+GTL+A
Sbjct: 36  MEEKLFNKHFVAITVINFIVYMVYYLFTVIIAFVATRELGAQTSQAGLATGIYILGTLLA 95

Query: 61  RLIFGKQLEVLGRKLVLRGGAIFYLLTTLAYFYMPSIGVMYLVRFLNGFGYGVVSTATNT 120
            RLIFGKQLEV GR+LVLRGGAIFYLLTTLAYFYMP+I +MYLVRFLNGFGYGVVSTATNT
Sbjct: 96  RLIFGKQLEVFGRRLVLRGGAIFYLLTTLAYFYMPTISMMYLVRFLNGFGYGVVSTATNT 155

Query: 121 IVTAYIPADKRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHINFKMVIVLCSILIAIVVL 180
            IVTAYIPA KRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHI+F+M+IVLCS+LI  VV+
Sbjct: 156 IVTAYIPARKRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHIDFRMIIVLCSVLIGCVVV 215

Query: 181 GAFVFPVKNITLNPEQLAKSKSWTIDSFIEKKAIFITIIAFLMGISYASVLGFQKLYTTE 240
            GAF FPVKN++LN EQLAK+KSWT+DSFIEKKA+FIT IAFLMGI+YASVLGFQKLYT+E
Sbjct: 216 GAFAFPVKNMSLNAEQLAKTKSWTVDSFIEKKALFITAIAFLMGIAYASVLGFQKLYTSE 275

Query: 241 INLMTVGAYFFIVYALVITLTRPSMGRLMDAKGDKWVLYPSYLFLTLGLALLGSAMGSVT 300
            I+L TVGAYFF+VYAL+IT+TRP+MGRLMDAKGDKWVLYPSYLFL +GL LLGS      +
Sbjct: 276 IHLTTVGAYFFVVYALIITITRPAMGRLMDAKGDKWVLYPSYLFLAMGLFLLGSVSSGGS 335

Query: 301 YLLSGALIGFGYGTFMSCGQAASIKGVEEHRFNTAMSTYMIGLDLGLGAGPYILGLVKDG 360
            YLLSGALIGFGYGTFMSCGQAASI+GV+EHRFNTAMSTYMIGLDLGLGAGPY+LGL+KD
Sbjct: 336 YLLSGALIGFGYGTFMSCGQAASIQGVDEHRFNTAMSTYMIGLDLGLGAGPYLLGLIKDL 395

Query: 361 FLGAGVQSFRELFWIAAIIPVVCGILYFLKS-SRQV 395
             LG+GV SFR LFW+AA+IP++C +LY LK+ +RQV
Sbjct: 396 ALGSGVASFRHLFWLAAVIPLICTLLYLLKTKTRQV 431
```

A related GBS gene <SEQ ID 8863> and protein <SEQ ID 8864> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      8.26
```

```
GvH: Signal Score (-7.5): -5.21
     Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 7 value: -14.33 threshold:  0.0
   INTEGRAL   Likelihood =-14.33   Transmembrane  167 - 183 ( 159 - 193)
   INTEGRAL   Likelihood = -7.96   Transmembrane   18 -  34 (  10 -  37)
   INTEGRAL   Likelihood = -7.75   Transmembrane  373 - 389 ( 369 - 392)
   INTEGRAL   Likelihood = -5.68   Transmembrane  214 - 230 ( 212 - 234)
   INTEGRAL   Likelihood = -4.78   Transmembrane  243 - 259 ( 241 - 262)
   INTEGRAL   Likelihood = -2.71   Transmembrane   48 -  64 (  47 -  65)
   INTEGRAL   Likelihood = -2.60   Transmembrane  283 - 299 ( 283 - 300)
   PERIPHERAL Likelihood =  0.69     341
modified ALOM score:   3.37

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01003(307 - 1494 of 1800)
EGAD|108032|BS3640(5   -   389   of   396)   hypothetical   protein   (Bacillus   subtilis)
GP|1684651|emb|CAB05383.1||Z82987 unknown similar to quinolon resistance protein NorA
(Bacillus subtilis) GP|2636170|emb|CAB15662.1||Z99122 similar to antibiotic resistance
protein (Bacillus subtilis) PIR|B70065|B70065 antibiotic resistance protein homolog ywoG -
Bacillus subtilis
%Match = 14.9
%Identity = 26.3  %Similarity = 53.4
Matches = 102  Mismatches = 178  Conservative Sub.s = 105

204       234       264       294       324       354       384       414
TTLTFVWAV*Y*HLYYTIEISYLLIFL*NVYENEIEKKEPFALEDKLFNKHFIGITILWFIVYMVYYLFTVIIAFIATKE
                                                  | :: | || : ::|: |:: :| |  ::     :|
                                                 MKKADAIWTKDFIMVLLVNLFVFVFFTFLTVLPIYTLQE
                                                             10        20        30        40

444       474       504       534       564       594       624       654
LGVSTSQAGLATGIYIVGTLIARLIFGKQLEVLGRKLVLRGGAIFYLLTTLAYFYMPSIGVMYLVRFLNGFGYGVVSTAT
|| : || ||   ::::  :| |   | :| :|:| :     :: |::: |  : :  ::  :||: |  : :::| |
LGGTESQGGLLISLFLLSAITTRPFSGATVERFGKKRMAIVSMAILFALSSFLYMPIHNFSLLLGLRFFQGIWFSIILTVT
          50        60        70        80        90       100       110       120

684       714       744       774       804       834       864       894
NTIVTAYIPADKRGEGINFYGLSTSLAAAIGPFVGTFMLDNLHINFKMVIVLCSILIAIVVLGAFVPFVKMITLNPEQLA
  |   ||| :|||: :: :| :|| |||||:|  ::    ::| :    ::::  :| :|:  |     :  :
GAIAADIIPAKRRGEGLGYFAMSMNLAMAIGPFLGLNLM--RVVSPFVFFTAFALFMVAGLLVSFLIKVPQSKDSGTTVF
        130       140       150       160       170       180       190

924       954       984       1014      1044      1074      1104      1134
KSKSWTIDSPIEKKAIFTIIAFLMGISYASVLGFQKLYTFTEINLMTVGAYFFIVYALVITLTRPSMGRLMDAKGDKWVL
:   :     || :|  : :::   |::|  :  ::   ::|  :  |||: :|: : : ||  |:| |  |    |:
R---FAFSDMFEKGALKIATVGLFISFCYSTVTSYLSVFAKSVDLSDISGYFFVCFAVTMMIARPFTGKLFDKVGPGIVI
          210       220       230       240       250       260       270

1164      1194      1224      1254      1284      1314      1344      1374
YPSYLFLTLGLALLGSANGSVTYLLSGNLIGFGYGTFMSCGQAASIKGVEERHFNTAMSTYMIGLDLGLGAGPYILGLVK
|||| | :::|| :|     :  |||||:|:|||: : | | :|:    ||  | :|:  :| |:  | |::||
YPSILIFSVGLCMLSFTHSGLMLLLSGHVIGLGYGSIVPCMQFLAIQKSPAHRSGFATATFPTFFDSGIAVGSYVFGL--
          290       300       310       320       330       340       350

1404      1434      1464      1494      1524      1554      1584      1614
DGFLGAGVQSPRELFWIAAIIPVVCSILYFLKSSRQVETKTI*KGGIKL*HKNMSVFLLLLMGLTSQNMR*KKG*MLLFV
          | ::  | :  :: :||     :  | :
----FVAAGAPSAIYLTAGLFVLIALLLYTWSQKKPAEAEGKVSIAE
          360       370       380       390
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1620

A DNA sequence (GBSx1715) was identified in *S.agalactiae* <SEQ ID 5001> which encodes the amino acid sequence <SEQ ID 5002>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0151(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06903 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 52/143 (36%), Positives = 84/143 (58%)

Query: 5   YERILLAIDGSYESELAVEKGINVALRNDAELLLTHVIDAHAYQSEGVFSDYVFDRQEQE 64
            Y  IL+A+DGS +++ A+ K  N A     A+L + HVID+ ++ +   +   V    E +
Sbjct: 2   YNHILVAVDGSTQAKRALYKAPNYAKEFKADLFICHVIDSRSFATVEQYDRTVVGAAELD 61

Query: 65  SADVLAYFEKLAHSKGLTKIKKITEIGNPKTLLAKDIPIREKADLIMVGATGLNTPERLL 124
              +L  + + A   G+ K+  I + G+PK  ++K I  +  DLI+ GATGLN   ER L
Sbjct: 62  GKKLLQRYSEEAEKAGVDKVETILDFGSPKANISKTTAQKVDIDLIITGATGLNAVERFL 121

Query: 125 IGSTSEYILRHSKVDMLVVRDSK 147
            +GS SE + RH+K D+L+VR+ +
Sbjct: 122 MGSVSESVARHAKCDVLIVRNKQ 144
```

There is also homology to SEQ ID 3658:

```
Identities = 105/150 (70%), Positives = 121/150 (80%)

Query: 1   MTQKYERILIAIDGSYESKLAVEKGINVALRNDAELLLTHVIDAHAYQSEGVPSDYVFDR 60
            M+ KY+RIL+AIDGSYESELA  KG+NVALRNDA LLL HVID  A QS   F  Y++++
Sbjct: 31  MSLKYKRILVAIDGSYESELAPNKGVNVALRNDATLLLVHVIDTRALQSVATFDTYIYEK 90

Query: 61  QEQESADVLAYFEKLAHSKGLTKIKKITSIGNPKTLLAKDIPIREKADLIMVGATGLNTP 120
             EQE+ DVL  FEK A   G+T IK+I E GNPK  LLA DIP RE ADLIMVGATGLNTF
Sbjct: 91  LEQEAKDVLDDFEKQAQIAGITNIKQIIEFGNPKNLLAHDIPDRENADLIMVGATGLNTF 150

Query: 121 ERLLIGSTSEYILRHSKVDMLVVRDSKKTL 150
            ERLLIGS+SEYI+RH+K+D+LVVRDS KTL
Sbjct: 151 ERLLIGSSSEYIMRHAKIDLLVVRDSTKTL 180
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1621

A DNA sequence (GBSx1716) was identified in *S.agalactiae* <SEQ ID 5003> which encodes the amino acid sequence <SEQ ID 5004>. This protein is predicted to be glycerol uptake facilitator protein (glpF). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.65   Transmembrane  261 - 277 ( 257 - 281)
   INTEGRAL    Likelihood = -5.73   Transmembrane  201 - 217 ( 199 - 222)
```

```
INTEGRAL     Likelihood = -4.51   Transmembrane    92 -  108 (  91 -  110)
INTEGRAL     Likelihood = -4.30   Transmembrane    44 -   60 (  42 -   62)
INTEGRAL     Likelihood = -2.18   Transmembrane    15 -   31 (  11 -   31)
INTEGRAL     Likelihood = -1.54   Transmembrane   150 -  166 ( 149 -  166)

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25231 GB:M58315 putative [Lactococcus lactis]
 Identities = 183/290 (63%), Positives = 228/290 (78%), Gaps = 10/290 (3%)

Query: 2   IEITWTVKYITEFIATAFLIILGNGAVANVDLKGTKGNNSGWIIIAIGYGLGVNMPALMF 61
            +++TWTVKYITEF+ TA LII+GNGAVANV+LKGTK +   W+II  GYGLGVM+PA+ F
Sbjct: 1   MDVTWTVKYITEFVGTALLIIMGNGAVANVELKGTKAHAQSWMIIGWGYGLGVNLPAVAF 60

Query: 62  GNVSGNHINPAFTLGLAPSGLFPWAHVGQYILAQILGAMFGQLVVVMVYQPYFVKTENPN 121
            GN++ + INPAFTLGLA SGLFPWAHV QYI+AQ+LGAMFGQL++VMVY+PY++KT+NPN
Sbjct: 61  GNIT-SQINPAFTLGLAASGLFPWAHVAQYIIAQVLGAMFGQLLIVMVYRPYYLKTQNPN 119

Query: 122 HVLGSFSTISALLDGQKSSKKAAVINGFLNEFVGSFVLFFGALALTKNYFGVE----LVG 177
                +LG+FSTI  +DD  + +R  A INGFLNEF+GSFVLFFGAAA T  +FG +    +
Sbjct: 120 AILGTFSTIDNVLDNSEKTRLGATINGFLNEFLGSFVLFFGAVAATNIPFGSQSITWMTN 179

Query: 178 KLVQAGYDQTTAATRISPYVTGSLA-----VAHLGIGFLVMTLVASLGGPTGPALNPARD 232
             L  G D +++      +V  S A      +AHL +GFLVM LV +LGGPTGP LNPARD
Sbjct: 180 YLKGQGADVSSSDVMNQIWVQASGASASKMIAHLFLGFLVMGLVVALGGPTGPGLNPARD 239

Query: 233 LGPRIVHRLLPKQILGQAKEDSKWWYAWVPVLAPIVASILAVALFKLLYL 282
             GPR+VH LLPK +LG+AK  SKWWYAWVPVLAPI+AS+ AVALFK++YL
Sbjct: 240 FGPRLVHSLLPKSVLGEAKGSSKWWYAWVPVLAPILASLAAVALFKMIYL 289
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5005> which encodes the amino acid sequence <SEQ ID 5006>. Analysis of this protein sequence reveals the following:

```
      Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL     Likelihood = -9.18   Transmembrane   293 -  309 ( 288 -  314)
INTEGRAL     Likelihood = -7.43   Transmembrane     2 -   18 (   1 -   20)
INTEGRAL     Likelihood = -7.38   Transmembrane   233 -  249 ( 228 -  256)
INTEGRAL     Likelihood = -5.57   Transmembrane   124 -  140 ( 123 -  142)
INTEGRAL     Likelihood = -2.87   Transmembrane    76 -   92 (  75 -   93)
INTEGRAL     Likelihood = -2.18   Transmembrane    47 -   63 (  43 -   63)
INTEGRAL     Likelihood = -1.54   Transmembrane   182 -  198 ( 181 -  198)

----- Final Results -----
               bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAA25231 GB:M58315 putative [Lactococcus lactis]
 Identities = 176/290 (60%), Positives = 228/290 (77%), Gaps = 10/290 (3%)

Query: 34  MEMTWTVKYITEFIATAFLIILGNGAVANVDLKGTKGHNSGWLVIAFGYGLGVNMPALMF 93
            M++TWTVKYITEF+ TA LII+GNGAVANV+LKGTK H   W++I +GYGLGVM+PA+ F
Sbjct: 1   MDVTWTVKYITEFVGTALLIIMGNGAVANVELKGTKAHAQSWMIIGWGYGLGVNLPAVAF 60

Query: 94  GNVSGNHINPAFTVGLAVSGLFPWAHVLQYVVAQLLGAIFGQLVVVMVYKPYFMCTENPN 153
            GN++ + INPAFT+GLA SGLFPWAHV QY++AQ+LGA+FGQLV++VMVY+PY++KT+NPN
Sbjct: 61  GNIT-SQINPAFTLGLAASGLFPWAHVAQYIIAQVLGAMFGQLLIVMVYRPYYLKTQNPN 119

Query: 154 HVLGSFSTISSLDNGQKDSHKASYINGFLNEFVGSFVLFFGALAITKNYFGVELVGKLIE 213
```

```
              +LG+FSTI ++D+  + +   + INGFLNEF+GSFVLFFGA+A T  +FG + +  +
Sbjct: 120 AILGTFSTIDNVDDNSEKTRLGATINGFLNEFLGSFVLFFGAVAATNIFFGSQSTTWMTN 179

Query: 214 ------AGYQTTAATQISPYVTGSLA---VAHIGIGFLVMVLVTSLGGPTGPALNPARD 264
              A    +   QI    +G+ A   +AH+ +GFLVM LV +LGGPTGP LNPARD
Sbjct: 180 YLKGQGADVSSSDVMNQIWVQASGASASKMIAHLFLGFLVMGLVVALGGPTGPGLNPARD 239

Query: 265 FGPRLLHHFLPKSVLGQAKGDSKWWYAWVPVVAPILAAIVAVAAFKYLYI 314
              FGPRL+H  LPKSVLG+AKG SKWWYAWVPV+APILA++ AVA FK +Y+
Sbjct: 240 FGPRLVHSLLPKSVLGEAKGSSKWWYAWVPVLAPILASLAAVALFKMIYL 289
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 240/281 (85%), Positives = 267/281 (94%)

Query: 2   IRITWTVKYITKFIATAFLIILGWGAVANVDLKGTKGNNSGWIIIAIGYGLGVMVPALMF 61
           +E+TWTVKYITEFIATAFLIILGNGAVANVDLKGTKG+NSGW++IA GYGLGVMNPALMF
Sbjct: 34  MEMTWTVKYITEFIATAFLIILGNGAVANVDLKGTKGHNSGWLVIAFGYGLGVMNPALMF 93

Query: 62  GNVSGNHINPAFTLGLAFSGLFPWAHVGQYILAQILGAMFGQLVVVMVYQPYFVKTENPN 121
           GNVSGNHINPAFT+GLA SGLFPWAHV QY++AQ+LGA+FGQLVVVMVY+PYF+KTENPN
Sbjct: 94  GNVSGNHINPAFTVGLAVSGLFPWAHVLQYVVAQLLGAIFGQLVVVMVYKPYFMKTENPN 153

Query: 122 HVLGSFSTISALDDGQKSSRKAAYINGFLNEFVGSFVLFFGALALTKNYPGVELVGKLVQ 181
           HVLGSFSTIS+LD+GQK S KA+YINGFLNEFVGSFVLFFGALALTKNYPGVELVGKL++
Sbjct: 154 HVLGSFSTISSLDNGQKDSHKASYINGFLNEFVGSFVLFFGALALTKNYPGVELVGKLIE 213

Query: 182 AGYDQTTAATRISPYVTGSLAVAHLGIGFLVMTLVASLGGPTGPALNPARDLGPRIVHRL 241
           AGYDQTTAAT+ISPYVTGSLAVAH+GIGFLVM LV SLGGPTGPALNPARD GPR++H
Sbjct: 214 AGYDQTTAATQISPYVTGSLAVAHIGIGFLVMVLVTSLGGPTGPALNPARDFGPRLLHHF 273

Query: 242 LPKQILGQAKEDSKWWYAWVPVLAPIVASILAVALFKLLYL 282
           LPK +LGQAK DSKWWYAWVPV+API+A+I +AVA FK LY+
Sbjct: 274 LPKSVLGQAKGDSKWWYAWVPVVAPILAAIVAVAAFKYLYI 314
```

A related GBS gene <SEQ ID 8865> and protein <SEQ ID 8866> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      2.81
GvH: Signal Score (-7.5): -3.6
     Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 6 value:  -8.65 threshold:  0.0
   INTEGRAL    Likelihood = -8.65   Transmembrane  261 - 277 ( 257 - 281)
   INTEGRAL    Likelihood = -5.73   Transmembrane  201 - 217 ( 199 - 222)
   INTEGRAL    Likelihood = -4.51   Transmembrane   92 - 108 (  91 - 110)
   INTEGRAL    Likelihood = -4.30   Transmembrane   44 -  60 (  42 -  62)
   INTEGRAL    Likelihood = -2.18   Transmembrane   15 -  31 (  11 -  31)
   INTEGRAL    Likelihood = -1.54   Transmembrane  150 - 166 ( 149 - 166)
   PERIPHERAL  Likelihood =  2.92       72
modified ALOM score:   2.23

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01006(304 - 1146 of 1446)
EGAD|14239|14211(1 - 289 of 289) hypothetical 30.9 kd protein in pepx 5'region (Lactococcus
lactis) SP|P22094|YDP1_LACLC HYPOTHETICAL 30.9 KDA PROTEIN IN PEPX 5'REGION (ORF1).
GP|455286|gb|AAA25206.1||M35865    ORF1    (put.);    putative    (lactococcus    lactis)
GP|149527|gb|AAA25231.1||M58315    putative    (lactococcus    lactis)    PIR|B43747|B43747
```

```
hypothetical protein (pepXP 5' region) - Lactococcus lactis subsp. cremoris
PIR|B43748|B43748 hypothetical protein (pepX 5' region) - Lactococcus lactis subsp. lactis
%Match = 37.5
%Identity = 64.4  %Similarity = 81.3
Matches = 183  Mismatches = 49  Conservative Sub.s = 48

123       153       183       213       243       273       303       333
*YASRS***ENLIN*IK*STR*SEPSTLFFIKYIWLKILLILFCDKLYNIKLTW*NG*CCKVFPGRKQGLIEITWTVKYI
                                                                    :::|||||||
                                                                    MDVTWTVKYI
                                                                          10

363       393       423       453       483       513       543       573
TEFIATAFLIILGNGAVANVDLKGTKGNNSGWIIIAIGYGLGVMMPALMPGNVSGNHINPAFTLGLAFSGLFPWAHVQQY
|||: ||:|||:|||||||:|||| :   |:|  |||||||:||: |||:: ::|||||||| |||||||| ||
TEFVGTALLIIMGNGAVANVELKGTKAHAQSWMIIGWGYGLGVMLPAVAPGNIT-SQINPAFTLGLAASGLFPWAHVAQY
          20        30        40        50        60        70        80

603       633       663       693       723       753       783       813
ILAQILGAMFSQLVVVMVYQPYFVKTENPNHVLGSFSTISALDDGQKSSRKAAYINGFLNEFVGSSFVLFFGALALTKNYF
|:|||||||||::|||:|::||:|||:|||:|||  :||  : :|  | |||||||:|||||||:| |  :|
ITAQVLGAMFSQLLIVMVYRPYYLKTQNPNAILGTFSTIDNVDDNSEKTRLGATINGFLNEFLGSFVLFFGAVAATNIFF
          100       110       120       130       140       150       160

831       861       885       906       936       966       996       1026
G----VSLVGKLVQAGYDQTTA--ATRISFVVTG---SLAVAHLGIGFLVMTLVASLGGPTGPALNPARDLGPRIVHRLL
|     :   |   | | :::   :|    :|   |  :||| :|||| || :|||||| |||||:|||:|| ||
GSQSITWMTNYLKGQGADVSSDVMNQIWVQASGASASKMIAHLFLGFLVMGLVVALGGPTGPGINPARDFGPRLVHSLL
          180       190       200       210       220       230       240

1056      1086      1116      1146      1176      1206      1236      1266
PKQILSQAKEDSKWYAWVPVLAPIVASILAVALFKLLYL**LKKDRFTGLFLF*I*KSASLAS*FRLWMTFGHSFFKGR
|| :||:|| ||||||||||||||:||: ||||||::||
PKSVLGEAKGSSKWYAWVPVLAPILASLAAVALFKMIYL
          260       270       280
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1622

A DNA sequence (GBSx1717) was identified in *S.agalactiae* <SEQ ID 5007> which encodes the amino acid sequence <SEQ ID 5008>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood = -8.70   Transmembrane  266 - 282 ( 262 - 290)
   INTEGRAL     Likelihood = -7.96   Transmembrane   25 -  41 (  24 -  50)
   INTEGRAL     Likelihood = -6.42   Transmembrane  110 - 126 ( 105 - 140)
   INTEGRAL     Likelihood = -6.26   Transmembrane  194 - 210 ( 190 - 215)
   INTEGRAL     Likelihood = -5.47   Transmembrane  290 - 306 ( 289 - 310)
   INTEGRAL     Likelihood = -4.35   Transmembrane  128 - 144 ( 127 - 147)
   INTEGRAL     Likelihood = -3.29   Transmembrane  157 - 173 ( 156 - 174)
   INTEGRAL     Likelihood = -2.76   Transmembrane  221 - 237 ( 221 - 240)

----- Final Results -----
              bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related sequence was also identified in GAS <SEQ ID 9177> which encodes the amino acid sequence <SEQ ID 9178>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 21
>>> Seems to have a cleavable N-term signal seq.
     INTEGRAL    Likelihood =-10.77   Transmembrane  139 - 155 ( 133 - 161)
     INTEGRAL    Likelihood = -8.28   Transmembrane  245 - 261 ( 240 - 269)
     INTEGRAL    Likelihood = -7.48   Transmembrane  269 - 285 ( 263 - 289)
     INTEGRAL    Likelihood = -7.06   Transmembrane   97 - 113 (  83 - 125)
     INTEGRAL    Likelihood = -6.10   Transmembrane  173 - 189 ( 169 - 194)
     INTEGRAL    Likelihood = -1.44   Transmembrane  200 - 216 ( 200 - 217)

----- Final Results -----
               bacterial membrane --- Certainty=0.531(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 225/301 (74%), Positives = 263/301 (86%)

Query: 10  LTVSLFFCRLDIMMETLLLHGIQLILIAMIITFYQIVRHIRSQKINPFKRFFTGLWIGF 69
           LT  +FFC+L  MME L+L  IQ +L+ AM+  F+ +V+H++  KINPFKRF+TG WIG
Sbjct: 1   LTAKVFFCKLVFMNEMLILRLIQALLVSAMLFIFFMLVKHLKKNKINPFKRFWTGFWIGL 60

Query: 70  VTDALDTLGIGSFATTTTFFKLTKLVKDDRKIPATWTAAHVLPVLLQSLCFIFVVKVEAL 129
           +TDALDTLGIGSFATTTT FKLTKLV DDR++P TMT AHVLPVL+QSLCFIFVVKVE L
Sbjct: 61  LTDALDTLGIGSFATTTTCFKLTKLVTDDRQLPGTMTVAHVLPVLIQSLCFIFVVKVEVL 120

Query: 130 TLITMAGAAFIGAFVGAKMTKNWHAPTVQRILGTLLITAAIIMLVRMITNPGAGISDSVH 189
           TL+ MA AAFIGA+ G  +TKNWHAPTVQRILG+LLI AAIIM+ R+I +PG  +SD++H
Sbjct: 121 TLLAMAAAAFIGAYFGTHITKNWHAPTVQRILGSLLIIAAIIMIIRIIYHPGEHLSDTIH 180

Query: 190 GLHGIWLFVGIGFNFIIGVLMTMGLGNYAPELIFFSLMGLSPAVAMPVMMLDAAMIMTAS 249
           GLHGIWLFVGIGFNFI+GVLMTMGLGNYAPELIFFSLMGLSP VAMPVMMLDAAMIMTAS
Sbjct: 181 GLHGIWLFVGIGFNFIVGVLMTMGLGNYAPELIFFSLMGLSPTVAMPVMMLDAAMIMTAS 240

Query: 250 STQFIKSGRVNWNGFAGLVTGGILGVIVAVLFLTNLDLNSLKTLVVGIVLFTGAMLIRSSF 309
           S+QFIK+ RV+W+GFAG+V+GGI+GV++AV FLTNLD+NSLK LV+ IV FTG MLIRSSF
Sbjct: 241 SSQFIKANRVSWDGFAGIVSGGIIGVLLAVFFLTNLDINSLKLLVIAIVFFTGGMLIRSSF 301
```

A related GBS gene <SEQ ID 8867> and protein <SEQ ID 8868> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:       2.32
GvH: Signal Score (-7.5): -5.59
     Possible site: 44
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 8 value:  -8.70 threshold:  0.0
   INTEGRAL    Likelihood = -8.70   Transmembrane  266 - 282 ( 262 - 290)
   INTEGRAL    Likelihood = -7.96   Transmembrane   25 -  41 (  24 -  50)
   INTEGRAL    Likelihood = -6.42   Transmembrane  110 - 126 ( 105 - 140)
   INTEGRAL    Likelihood = -6.26   Transmembrane  194 - 210 ( 190 - 215)
   INTEGRAL    Likelihood = -5.47   Transmembrane  290 - 306 ( 289 - 310)
   INTEGRAL    Likelihood = -4.35   Transmembrane  128 - 144 ( 127 - 147)
   INTEGRAL    Likelihood = -3.29   Transmembrane  157 - 173 ( 156 - 174)
   INTEGRAL    Likelihood = -2.76   Transmembrane  221 - 237 ( 221 - 240)
   PERIPHERAL  Likelihood =  3.87       67
modified ALOM score:   2.24

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5009> which encodes amino acid sequence <SEQ ID 5010>:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.77   Transmembrane  151 - 167 ( 145 - 173)
   INTEGRAL    Likelihood = -9.13   Transmembrane   22 -  38 (  15 -  42)
   INTEGRAL    Likelihood = -8.28   Transmembrane  257 - 273 ( 252 - 281)
   INTEGRAL    Likelihood = -7.48   Transmembrane  281 - 297 ( 275 - 301)
   INTEGRAL    Likelihood = -7.06   Transmembrane  109 - 125 (  95 - 137)
   INTEGRAL    Likelihood = -6.10   Transmembrane  185 - 201 ( 181 - 206)
   INTEGRAL    Likelihood = -1.44   Transmembrane  212 - 228 ( 212 - 229)
   INTEGRAL    Likelihood = -0.27   Transmembrane    5 -  21 (   5 -  21)

----- Final Results -----
               bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS sequences follows:

```
 Score =  405 bits (1029), Expect = e-115
 Identities = 198/301 (65%), Positives = 228/301 (74%)

Query: 1   LTAKVFPCKLVFMNEMLLILRLIQALLVSAMLFIFFMLVKHLKKNKINPFKRFWTGFWIGL 60
           LT  +FFC+L  MNE L+L  IQ +L+ AM+  F+ +V+H++  KINPFKRF+TG WIG
Sbjct: 10  LTVSLFFCRLDIMNETLLLHGIQLILIIAMIITFYQIVRHIRSQKINPFKRFFTGLMIGF 69

Query: 61  LTDALDTLGIGSFATTTTCFKLTKLVTDDRQLPGTMTVAHVLFVLIQSLCPIFVVKEVEVX 120
           +TDALDTLGIGSFATTTT FKLTKLV DDR++P TMT AHVLFVL+QSLCPIFVVKEVE
Sbjct: 70  VTDALDTLGIGSFATTTTFFKLTKLVEDDRKIPATMTAAHVLFVLLQSLCPIFVVKEVEAL 129

Query: 121 XXXXXXXXXFIGAYFGTHITGNWHAPTVQRILGSLLXXXXXXXXXXXXYHPGHLSDTIH 180
                FIGA+ G  +TGNWHAPTVQRILG+LL              +PG  +SD++H
Sbjct: 130 TLITMAGAAFIGAFVGAKMTKNWHAPTVQRILGTLLITAAIIMLYRMITNPGAGISDSVH 189

Query: 181 GLHGIWLFVGIGFNFIVGVLMTMGLGNYAPELIFFSLMGLSPTVAMPVMMLDAAMIMTAS 240
           GLHGIWLFVGIGFNFI+GVLMTMGLGNYAPELIFFSLMGLSP VAMPVMMLDAAMIMTAS
Sbjct: 190 GLHGIWLFVGIGFNFIIGVLMTMGLGNYAPELIFFSLMGLSPAVAMPVMMLDAAMIMTAS 249

Query: 241 SSQPIKANRVSWDXXXXXXXXXXXXXXXXXFPLTNLDINSLKLLVIAIVFTFGGMLIRSSF 301
           S+QFIK+ RV+W+                 FLTNLD+NSLK LV+ IV FTG MLIRSSF
Sbjct: 250 STQFIKSGRVMWNGFAGLVTGGILGVIVAVLFLTNLDLNSLKTLVVGIVLFTGAMLIRSSF 310
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1623

A DNA sequence (GBSx1718) was identified in *S.agalactiae* <SEQ ID 5011> which encodes the amino acid sequence <SEQ ID 5012>. This protein is predicted to be C3-degrading proteinase. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2851(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD37110 GB:AF112358 C3-degrading proteinase [Streptococcus pneumoniae]
```

```
 Identities = 92/240 (38%), Positives = 142/240 (58%), Gaps = 11/240 (4%)

Query: 12  PVLRVNNRDLNIAFYQESLGFKLISEENAIAVFSAWQNKEASFIIEESPTYRTRAVNGTK 71
           P L+ NNR LN  FY E+LG K + EE+A          E   ++EE+P+ RTR V G K
Sbjct: 11  PTLKANNRKLNETFYIETLGMKALLEESAFLSLGDQTGLE-KLVLESAPSMRTRKVEGRK 69

Query: 72  KLAKIIVKSQDAKDIEKLLANGAQAIQVYQGQNGYAYETVSPEGDLFLLHAEDDLSQLVA 131
           KLA++IVK ++  +IE +L+      ++Y+GQNGYA+E  SPE DL L+HAEDD++ LV
Sbjct: 70  KLARLIVKVENPLEIEGILSKTDSIHRLYKGQNGYAFEISPEDDLILIHAEDDIASLVE 129

Query: 132 I-ERPELEKKDDTTGLSNFAFQGISLNVPDAVKAEAFYDKVFAGKFPINLSFKEAQQQDL 190
           + E+PE +    +  LS F   S+ L++P  +  E+F +   + +  +L F  AQSQDL
Sbjct: 130 VGEKPEFQTDLASISLSKPEI-SMELHLPTDI--ESFLE---SSEIGASLDFIPAQQQDL 183

Query: 191 QIAPNETWDIEILECCVNKDINLNDLKSTPESLGILDVYLDSKEKILVISDTSNIEIWISK 250
            +   TWD+ +L+  VNE  ++  L+  PES   + ++   EK  +  D +N+E+W  +
Sbjct: 184 TVDNTVTWDLSMLKFLVNE-LDIASLRQKPES--TEYPIPKSEKPFLGKDRNNVELWFEE 240
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5013> which encodes the amino acid sequence <SEQ ID 5014>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3267(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 130/250 (52%), Positives = 177/250 (70%)

Query: 1   MTLFHSLTFKEPVLRVNNRDLNIAFYQESLGFKLISEENAIAVFSAWQNKEASFIIEESP 60
           MTL  ++TFK PVLRVN+RDLNIAFYQ +LG+LSEENAIA+FS+W   +  F+IEESP
Sbjct: 1   MTLMENITFKTPVLRVNDRDLNIAFYQNNLGIRLVSEENAIAIFSSWGEQECFVIEESP 60

Query: 61  TYRTRAVNGTKKLAKIIVKSQDAKDIEKLLANGAQAIQVYQGQNGYAYETVSPEGDLFLL 120
           + RTRAV G KK+  I++K+   K+IE+LLA+GA     +++GQNGYA+ET+SPEGD FLL
Sbjct: 61  SVRTRAVEGPKKVNTIVIKTNQPKEIEQLLAHGAHYDALFKGQNGYAFETISPEGDRFLL 120

Query: 121 HAEDDLSQLVAISRPELEKKDDTTGLSNFAFQSISLNVPDAVKAEAFYDKVFAGKFPINL 180
           HAE D+  L   + P LEK     GL+ F F  I LNV    +++AFY  +F+ + PI +
Sbjct: 121 HAEQDIKHLQGTDLPSLEKDATFKGLTQFKFDIIVLNVISEERSKAFYRDLFSDQLPITN 180

Query: 181 SFKEAQQQDLQIAPNETWDIEILECCVNEDTNLNDLKSTPESLGLDVVLDSKEKLLVISD 240
            F + +G DL I  F+  WD+EILE  V++D ++  LK+T E G  VY+D K K+LV+SD
Sbjct: 181 DFIQEEGPDLAIDPHIAWDLEILEFQVSKDYDMKVLKATLEEDGHKVIIDKKHKVLVLSD 240

Query: 241 TSNIEIWISK 250
            S IE+W +K
Sbjct: 241 PSQIEVWFTK 250
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1624

A DNA sequence (GBSx1719) was identified in *S.agalactiae* <SEQ ID 5015> which encodes the amino acid sequence <SEQ ID 5016>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
bacterial cytoplasm --- Certainty=0.2510(Affirmative) < Succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < Succ>
bacterial outside --- Certainty=0.0000(Not Clear) < Succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAC16441 GB:AL450165 putative esterase [Streptomyces coelicolor]
 Identities = 89/323 (27%), Positives = 143/323 (43%), Gaps = 51/323 (15%)

Query: 10  NTVLELIKEQIKDNLYHGASLAIY-ENGEWHEHYLGT-------IDGNEKVKAGLVYDLA 61
           +T+ EL+ E  +  + GA+ ++     G     + GT          +GG++      V+DLA
Sbjct: 2   STLAELLAEGRRQRICSGAAWSVGGPQGPLDRGWTGTRCWDGPPLDGDD------VWDLA 55

Query: 62  SVSKVVGVGTLLAKLVVQGTIDIDKPLRYYYPTFH---HQTLTVRQLATHSSGIDPFIP- 117
           SV+K +  G ++  LV +G + +D  +  Y P +         LTVRQL  H+SGI    +P
Sbjct: 56  SVTKPIA-GLVVMALVERGALGLDDTVGGYLPDYRGDGKAELTVRQLLAHTSGIPSQVPL 114

Query: 118 NRDQLNATQLKDAINHIKVLEDKSFK--YTDINFLLLGFMLEEVLGDSLDKLFKRYIFTP 175
            RD       L +A+  + +       +  Y+    F++LG + E    G+ L+ L +R +  P
Sbjct: 115 YRDHPTRAALLEAVRLLPLTAQPGTRVQYSSQGFIVLGLIAEAAAGEPLEALVERLVCAP 174

Query: 176 FQMKETSFGPRVEAVPTVVGIND---------GIVHDPKAKVLGKHTGSAGLFSTIDDLQ 226
           +++T F P       V   D              G VHD  A VLG    G AGLFST+ D++
Sbjct: 175 LGLRDTVFRPDAGRRARAVATEDCPWRGRRVVGEVHDENAVVLGGVGSHAGLFSTLADMS 234

Query: 227 RFSIHYL--------KDDFA-KPLWNNYSLSKSRSLAWD------------IDKDWINHT 265
           R              + FA     +   L+  R+LAW                +   HT
Sbjct: 235 RLGAALAAGGRGLLRPETFALMTAAHTDGLALERALAWQGRDPVGSPAGEVPGPESYGHT 294

Query: 266 GYTGPFIALNYQKQAAAIFLTNR 288
           G+TG  + ++   + A+ LTNR
Sbjct: 295 GFTGSLWVDPATRRYAVLLTNR 317
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3885> which encodes the amino acid sequence <SEQ ID 3886>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.33   Transmembrane   57 -  73 (  57 -  74)

----- Final Results -----
               bacterial membrane --- Certainty=0.1532(Affirmative) < Succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < Succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < Succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 174/302 (57%), Positives = 229/302 (75%), Gaps = 1/302 (0%)

Query: 9   TNTVLELIKEQIKDNLYHGASLAIYENGEWHEHYLGTIDGNEKVKAGLVYDLASVSKVVG 68
           T  V++ I+  +   +Y GASLA++++G W E+++GTIDG   V A LVYDLASVSKVVG
Sbjct: 6   TLAVIKCIENLHKKVYKGASLALFQSGRWQEYHIGTIDGRRPVDANLVYDLASVSKVVG 65

Query: 69  VCTLLAKLVYQGTIDIDKPLRYYYPTFHHQTLTVRQLATHSSGIDPFIPNRDQLNATQLK 128
           V T+    L+  GT+ +D PL+ YYP+     T+T+RQL TH+SG+DP+IPNRD LNA QL+
Sbjct: 66  VATICNILLNNGTLALDDPLKVYYPSIADATVTIRQLLTHTSGLDPYIPNRDVLNAQQLR 125

Query: 129 DAINHIKVLEDKSFKYTDINFLLLGFMLEEVLGDSLDKLFKKYIFTPFQMKETSFGPRVE 188
           +A+NH+    E+K+F YTD+NFLLLGFMLEE+   +SLD++F + IFTPF M  TSFGPR E
Sbjct: 126 KALNHLTQKRNKNFYYTDVNFLLLGFMLERLFSESLDQIFDKTIFTPFGNYHTSFGPRPE 185

Query: 189 AVPTVVGINDGIVHDPKAKVLGKHTGSAGLFSTIDDLQRFSIHYLKDDFAKPLWNNYSLS 248
           AVPT+ G++DG VHDPKAK+L KH+GSAGLFST+ DL+  FS HYL D F+  LW NYS
Sbjct: 186 AVPTLKGVSDGEVHDPKAKILKKHSGSAGLFSTLADLESFSNHYLNDPFSDCLWRNYSQQ 245

Query: 249 K-SRSLAWDIDKDWINHTGYTGPFIALNYQKQAAAIFLTNRTFPSYDDRPLWIKKRRHVQE 307
              RSL W++D DWI+HTGYTGPF+ LN ++Q AAIFLTNRT+  DD+  W+K+R+ +
```

```
Sbjct: 246 TIERSLGWNLDGDWISHTGYTGPFLMLNKKEQTAAIFLTNRTYDEDDKSKWLKERQLLYN 305

Query: 308 AI 309
           A+
Sbjct: 306 AL 307
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1625

A DNA sequence (GBSx1720) was identified in *S.agalactiae* <SEQ ID 5017> which encodes the amino acid sequence <SEQ ID 5018>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0935(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA25177 GB:D21804 FMN-binding protein [Desulfovibrio vulgaris]
 Identities = 53/124 (42%), Positives = 76/124 (60%), Gaps = 2/124 (1%)

Query: 1   MLNHKFLQVLKYEGVVSITSWIELAPHVTNTWNSYLTITDDQRILAPAAGMTHLENDLNN 60
           ML   F +VLK EGVV+I +  E  PH+ NTWNSYL + D  RI+ P  GM   E ++
Sbjct: 1   MLEGTFFEVLKNEGVVAIATQGEDGPHLVNTWNSYLKVLDGNRIVVPVGGMHKTEANVAR 60

Query: 61  NSKIIMTLGSREVEGRDGVQGTGFRIEGTAKLLEAGSDFEIVKEKYPFLRKVLEVTPINV 120
           + +++MTLGSR+V GR+G  GTGF I G+A     G +FE +  ++ + R  L +T ++
Sbjct: 61  DERVLMTLGSRKVAGRNG-PGTGFLIRGSAARTDGPEPEAI-ARFKWARAALVITVVSA 118

Query: 121 IQLL 124
           Q L
Sbjct: 119 EQTL 122
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1626

A DNA sequence (GBSx1721) was identified in *S.agalactiae* <SEQ ID 5019> which encodes the amino acid sequence <SEQ ID 5020>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3799(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1627

A DNA sequence (GBSx1722) was identified in *S.agalactiae* <SEQ ID 5021> which encodes the amino acid sequence <SEQ ID 5022>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3175(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10123> which encodes amino acid sequence <SEQ ID 10124> was also identified.

The protein has homology to a pyruvate formate-lyase from *S.mutans*:

```
>GP:BAA09085 GB:D50491 Pyruvate formate-lyase [Streptococcus mutans]
 Identities = 709/770 (92%), Positives = 750/770 (97%)

Query: 7   MATVKTNTDIFEQAWEGFKGVDWKEKASIARFVQANYAPYDGDESFLAGATERSLHIKKV 66
            MATVKTNTD+FE+AWEGFKG DWK++ASI+RFVQ NY PYDG ESFLAG TERSLHIKKV
Sbjct: 1   MATVKTNTDVFEKAWEGFKGTDWKDRASISRFVQDNYTPYDGGESFLAGPTERSLHIKKV 60

Query: 67  IEETKAHYEETRFPMDTRVASISELPAGFIDKDNELIPGIQNDELPKLNFMPKGGIRMAE 126
            +EETKAHYEETRFPMDTR+ SI+++PAG+IDK+NELIPGIQNDELPKLNFMPKGGIRMAE
Sbjct: 61  VEETKAHYEETRFPMDTRITSIADIPAGYIDKENELIPGIQNDELPKLNFMPKGGIRMAE 120

Query: 127 TTLKRMKYEPDPAVHEIFTKYATTVNDGIPRAYTSNIRRARHAHTVTGLPDAYSRGRIIG 186
            T LKE+GYEPDPAVHEIFTKYATTVNDGIPRAYTSNIRRARHAHTVTGLPDAYSRGRIIG
Sbjct: 121 TALKEHGYEPDPAVHEIFTKYATTVNDGIPRAYTSNIRRARHAHTVTGLPDAYSRGRIIG 180

Query: 187 VYARLAVYGADYLMQEKVNDWNALNDIDEESIRLREEINLQYQALGEVVKLGDLYGVDVR 246
            VYARLA+YGADYLMQEKVNDWN++ +IDEESIRLREEINLQYQALGEVV+LGDLYG+DVR
Sbjct: 181 VYARLADYGADYLMQEKVNDWNSIAEIDESSIRLREEINLQYQALGEVVRLGDLYGLDVR 240

Query: 247 KPAMNTKGAIQWVNIAPMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTSEIQEFV 306
            KPAMN KGAIQW+NIAPMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTSEIQEFV
Sbjct: 241 KPAMNVKEAIQWINIAPMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTSEIQEFV 300

Query: 307 DDFVLKLRTVKFARTKAYDALYSGDPTFITTSMAGMGADGRHRVTKMDYRFLNTLDNIGN 366
            DDFV+KLRTVKFARTKAYD LYSGDPTFITTSMAGMGADGRHRVTKMDYRFLNTLDNIGN
Sbjct: 301 DDFVMKLRTVKFARTKAYDELYSGDPTFITTSMAGMGADGRHRVTKMDYRFLNTLDNIGN 360

Query: 367 SPEPNLITVLWSDQLPYAFRRYCMSMSHKHSSIQYEGVSTMAREGYGHMSCISCCVSPLDP 426
            +PEPNLITVLWS +LPY+FR YCMSMSHKHSSIQYEGVTTMAREGYGEMSCISCCVSPLDP
Sbjct: 361 APEPNLITVLWSSKLPYSFRHYCMSMSHKHSSIQYEGVTTMAREGYGEMSCISCCVSPLDP 420

Query: 427 ENEDKRRNLQYFGARVNVMKALLTGLNGGYDDVHKDYKVFDIDPIRDEVLNFDTVKANFE 486
            ENED+R+NLQYFGARVNV+KALLTGLNGGYDDVHKDYKVFD++PIRDEVL+F+TVKANFE
Sbjct: 421 ENEDRRENLQYFGARVNVLKALLTGLNGGYDDVHKDYKVFDVEPIRDEVLDFETVKANFE 480

Query: 487 KSLDWLTDTYVDAMNIIHYMTDKYNYEAVQMAFLPSHVRANMGPGICGFANTVDSLSAIK 546
            K+LDWLTDTYVDAMNIIHYMTDKYNYEAVQMAFLP+ V+ANMGPGICGF+NTVDSLSAIK
Sbjct: 481 KALDWLTDTYVDAMNIIHYMTDKYNYEAVQMAFLPTRVKANMGPGICGFSNTVDSLSAIK 540

Query: 547 YATVKPIRDEDGYIYDYETWSDPFRYGEDDDRVDSIAEWLLEAFHRLAKHKLYKDAEAT 606
            YATVKPIRDEDGYIYDYETWG+PFRYGEDDDRVDSIAEWLLEAFH RLA+HKLYKD+EAT
Sbjct: 541 YATVKPIRDEDGYIYDYETWGNPFRYGEDDDRVDSIAEWLLEAFHTRLARHKLYKDSEAT 600

Query: 607 VSLLTITSNVAYSKQTGNSPVHKGVYLNSDGSVNLSKVEFPSPGANPSNKAKGGWLQNLW 666
```

```
              VSLLTITSNVAYSKQTGNSPVHKGVYLNEDGSVNLSKVEFFSPGANPSNKA GGWLQNLN
Sbjct: 601 VSLLTITSNVAYSKQTGNSPVHKGVYLNEDGSVNLSKVEFFSPGANPSNKASGGWLQNLN 660

Query: 667 SLSKLDFAHANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFPENGGQHVNLNVMDLKD 726
           SL KLDFAHANDGISLTTQVSP+ALGKTFDEQV NLVT+LDGYFP GGQHVNLNVMDLKD
Sbjct: 661 SLKKLDFAHANDGISLTTQVSPKALGKTFDEQVANLVTILDGYFPEGGQQHVNLNVMDLKD 720

Query: 727 VYDKIMNGEDVIVRISGYCVNTKYLTPEQKTELTQRVFHEVLSMDDALTN 776
           VYDKIMNGEDVIVRISGYCVNTKYLT EQKTELTQRVFHEVLSMDDA T+
Sbjct: 721 VYDKIMNGEDVIVRISGYCVNTKYLTKEQKTELTQRVFHEVLSMDDAATD 770
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5023> which encodes the amino acid sequence <SEQ ID 5024>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3184(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 701/773 (90%), Positives = 742/773 (95%), Gaps = 1/773 (0%)

Query: 2   FKEKTMATVKTNTDIFEQAWEGFKGVDWKEKASIARFVQANYAPYDGDESFLAGATERSL 61
           FKEK MATVKTNTD+FE+AWEGFKG DWKEKAS++RFVQANY PYDGDESFLAGATERSL
Sbjct: 5   FKEKFMATVKTNTDVFEKAWEGFKGTDWKEKASVSRFVQANYTPYDGDESFLAGATERSL 64

Query: 62  HIKKVIEETKAHYEETRFPMDTRVASISELPAGFIDKDNELIFGIQNDELPKLNFMPKGG 121
           HIKKVIEETKAHYE TRFP DTR  SI+++PAGFIDK+NELI+GIQNDELPKLNFMPKGG
Sbjct: 65  HIKKVIEETKAHYEATRFPVDMRPTSIADIPAGFIDKSNELIYGIQNDELPKLNFMPKGG 124

Query: 122 IRMAETTLKENGYEPDPAVHEIFTKYATTVNDGIPRAYTSNIRRAHAHTVTGLPDAYSR 181
           IRMAETTLKENGYEPDPAVHEIFTKY TTVNDGIPRAYTSNIRRAHAHTVTGLPDAYSR
Sbjct: 125 IRMAETTLKENGYEPDPAVHEIFTKYVTTVNDGIPRAYTSNIRRAHAHTVTGLPDAYSR 184

Query: 182 GRIIGVYARLAVYGADYLMQEKVNDWNALNDIDEBSIRLREEINLQYQALGEVVKLGDLY 241
           GRIIGVYARLA+YGADYLMQEKVNDWNA+ +IDEKSIRLREE+NLQYQALGEVVKLGDLY
Sbjct: 185 GRIIGVYARLALYGADYLMQEKVNDWNAITEIDEESIRLREEVNLQYQALGEVVKLGDLY 244

Query: 242 GVDVRKPAMNTKEAIQWVNIAFMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTESE 301
           GVDVR+PA N KEAIQWVNIAFMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTESE
Sbjct: 245 GVDVRRPAQNVKEAIQWVNIAFMAVCRVINGAATSLGRVPIVLDIFAERDLARGTFTESE 304

Query: 302 IQEFVDDFVLKLRTVKFARTKAYDALYSGDPTFITTSMAGMGNDGRHRVTKMDYRFLNYL 361
           IQEFVDDFVLKLRTVKF RTKAYDALYSGDPTFITTSMAGMG DGRHRVTKMDYRFLNYL
Sbjct: 305 IQEFVDDFVLKLRTVKFGRTKAYDALYSGDPTFITTSMAGMGNDGRHRVTKMDYRFLNYL 364

Query: 362 DNIGNSPEPNLTVLWSDQLPYAFRRYCMSMSHKHSSIQYEGVSTMAKEGYGEMSCISCCV 421
           DNIGNSPEPNLTVLW+DQLP  FRRYCM MSHKHSSIQYEGV+TMAKEGYGEMSCISCCV
Sbjct: 365 DNIGNSPEPNLTVLWTDQLPETFRRYCMRNMSHKHSSIQYEGVTTMAKEGYGEMSCISCCV 424

Query: 422 SPLDPENEDKRHNLQYPGARVNVMKALLTGLNGGYDDVHKDYKVPD-IDPIRDEVLNFDT 480
           SPLDPENE++RHN+QYPGARVNV+KALLTGLNGGYDDVH+DYKVP+ ++PI  EVL +D
Sbjct: 425 SPLDPENEQQRHNIQYPGARVNVLKALLTGLNGGYDDVHRDYKVPNVVEPITSKVLKYDE 484

Query: 481 VKANFEKSLDWLTDTYVDNMNIIHYMTDKYNYEAVQMAFLPSHVRANMGFGICGFANTVD 540
           V ANFEKSLDWLTDTYVDA+NIIHYMTDKYNYEAVQMAFLP+H RANMGFGICGFANTVD
Sbjct: 485 VMANFEKSLDWLTDTYVDALNIIHYMTDKYNYEAVQMAFLPTHQRANMGFGICGFANTVD 544

Query: 541 SLSAIKYATVKPIRDEDGYIYDYETVGDFPRYGEDDDRVDSIAEWLLSAFHGRLAKHKLY 600
           +LSAIKYATVK IRDE+GYIYDYE  GDFPRYGEDDDRVD IA+WL+EA+H RLA HKLY
Sbjct: 545 TLSAIKYATVKTIRDENGYIYDYEVTGDFPRYGEDDDRVDDIAKWLMEAYHTRLASHKLY 604

Query: 601 KDAEATVSLLTITSNVAYSKQTGNSPVHKGVYLNEDGSVNLSKVEFFSPGANPSNKAKSG 660
```

```
              K+ARA+VSLLTITSNVAYSKQTGNSPVH+GV+LNEDG+VN S+VEPFSPGANPSNKAKGG
Sbjct: 605 KNASASVSLLTITSNVAYSKQTGNSPVHRGVFLNEDGTVNTSQVEPFSPGANPSNKAKGG 664

Query: 661 WLQNLNSLSKLDFAHANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFENGGQHVNLN 720
           WLQNLNSL+KL+F+HANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFENGGQHVNLN
Sbjct: 665 WLQNLNSLAKLEFSHANDGISLTTQVSPRALGKTFDEQVDNLVTVLDGYFENGGQHVNLN 724

Query: 721 VMDLKDVYDKIMNGEDVIVRISGYCVNTKYLTPEQKTELTQRVFHEVLSMDDA 773
           VMDL DVYDKIMNGEDVIVRISGYCVNTKYLTPEQKTELTQRVFHEVLSMDDA
Sbjct: 725 VMDLNDVYDKIMNGEDVIVRISGYCVNTKYLTPEQKTELTQRVFHEVLSMDDA 777
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1628

A DNA sequence (GBSx1723) was identified in *S.agalactiae* <SEQ ID 5025> which encodes the amino acid sequence <SEQ ID 5026>. This protein is predicted to be DNA-damage inducible protein P (dinP). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10121> which encodes amino acid sequence <SEQ ID 10122> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF95431 GB:AE004300 DNA-damage-inducible protein P [Vibrio cholerae]
 Identities = 136/349 (38%), Positives = 210/349 (59%), Gaps = 14/349 (4%)

Query: 12  INDTSRKIIHIDMDAFFASVEERDNPSLKGKPVIIGSDPRKTGGRGVVSTCNYEARKFGV 71
           + D  RKIIH+DMD FFA+VE RDNP+ +    +G   ++    RGV+STCNY+ARKFGV
Sbjct: 1   MQDRIRKIIHVDMDCFFAAVEMRDNPAYREIALAVGGHEKQ---RGVISTCNYQARKFGV 57

Query: 72  HSAMSSKEAYERCPQAIFISGNYQKYRQVGMEVRDIFKKYTDLVEPMSIDEAYLDVTENK 131
            SAM + +A + CPQ   + G    Y+ V  +++ IF++YT L+EP+S+DEAYLDV+E+
Sbjct: 58  RSAMPTAQALKLCPQLHVVPGRMSVYKSVSQQIQTIFQRYTSLIEPLSLDEAYLDVSEST 117

Query: 132 MGIKSAVKLAKMIQVDIKNDVHLTCSAGISYNKFLAKLASDFEKPKGLTLILPDQAQDFL 191
            SA  +A+ I+ DIW +++LT SAG++  KFLAK+ASD  KP GL ++ PD+ Q+ +
Sbjct: 118 AYQGSATLTAQAIRRDIWQEINLTASAGVAFIKFLAKVASDLNKPDGLYVVTPDKVQEMV 177

Query: 192 KPLPIEKFHGVGKRSVEKLEALGVYTGEDLLSLSEISLIDMFGRPGVDLYRKARGINASP 251
            LP+EK  GVGK ++EKLE  G+Y G D+        L+  FGR G  L++K+ GI+
Sbjct: 178 DSLPLKKIPGVGKVALEKLEQAGLYVGADVRADYRKLLHQFGRLGASLWKKSHGIDERE 237

Query: 252 VKPDRVRKSIGSEKTYGKLLYNEADIKAEISKNVQRVVASLEKNKKVGKTIV---LKVRY 308
           V  +R RKS+G E T+ + +    I + +  + +       + I+   +KV++
Sbjct: 238 VVTERERKSVGVEYTFSQNISTFQSCWQVIEQKLYPELDARLSRAHPQRGIIKQGIKVKF 297

Query: 309 ADFETLTKRMTLEEYTQDF--QIIDQVAKAIFDTLEESVFGIRLLGVTV 355
           ADF+  T          D+  ++++QV          +    IRLLG++V
Sbjct: 298 ADFQQTTIEHVHPALELDYFHELLEQV------LTRQQGREIRLLGLSV 340
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5027> which encodes the amino acid sequence <SEQ ID 5028>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.1921(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 276/363 (76%), Positives = 323/363 (88%)

Query: 6   MLIPPLINDTSRKIIHIDMDAFFAGVEERDNPSLGKPVIIGSDPRKTGGRGVVSTCNYR 65
           MLIPPLINDTSRKIIHIDMDAFFAAVEERDNPALGKPVVIGKDPRETGGRGVVSTCNYE
Sbjct: 1   MLIPPLINDTSRKIIHIDMDAFFAAVEERDMPALGKPVVIGKDPRETGGRGVVSTCNYE 60

Query: 66  ARKFGVHSAMSSKEAYERCPQAIFISGNYQKYRQVGMEVRDIFKKYTDLVEPMSIDEAYL 125
           ARK+G+HSAMSSKEAYERCP+AIFISGNY+KYR VG ++R IFK+YTD+VEPMSIDEAYL
Sbjct: 61  ARKYGIHSAMSSKEAYERCPKAIFISGNYEKYRTVGDQIRRIFKRYTDVVEPMSIDEAYL 120

Query: 126 DVTENKMGIKSAVKLAKMIQYDIWNDVHLTCSAGISYNKFLAKLASDFEKPKGLTLILPD 185
           DVT+NK+GIKSAVK+AK+IQ+DIW +V LTCSAG+SYNKFLAKLASDFEKP GLTL+L +,
Sbjct: 121 DVTDNKLGIKSAVKIAKLIQHDIWKSVGLYCSAGVSYNKFLAKLASDFEKPHGLTLVLKE 180

Query: 186 QAQDFLKPLPIEKFHGVGKSVEKLHALGVYTGEDLLSLSESISLIDMFGRYGVDLYRKAR 245
            A  FL   LPIEKFHGVGK+SV+KLH +G+YTG+DLL++ E++LID FGRPG+DLYRKAR
Sbjct: 181 DALCFLAKLPIEKFHGVGKKSVKKLEDNGIYTYGQDLLAVPEMTLIDHPGRPGPDLYRKAR 240

Query: 246 GINASPVKPDRVRKSIGSEKTYGKLLYNEADIKAEISKNVQRVVASLEKNKKVGKTIVLK 305
           GI+ SPVK DR+RKSIGSE+TY KLLY E DIKAEISKNV+RV A L+ +KK+GKTIVLK
Sbjct: 241 GISNSPVKYDRIRKSIGSERTYAKLLYQETDIKAEISKNVKRVAALLQDHKKLGKTIVLK 300

Query: 306 VRYADFTLTTKRMTLEEYTQDFQIIDQVAKAIPDTLEESVFGIRLLGVTVTTLENEHEAI 365
           VRYADF TLTKR+TL E T++   I+QVA  IPD+L E+  GIRLLGVT+T LE++   I
Sbjct: 301 VRYADFTTLTKRVTLPELTRNAAQIRQVAGDIFDSLSENPAGIRLLGVTMTNLEDKVADI 360

Query: 366 YLD 368
           LD
Sbjct: 361 SLD 363
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1629

A DNA sequence (GBSx1724) was identified in *S.agalactiae* <SEQ ID 5029> which encodes the amino acid sequence <SEQ ID 5030>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.11   Transmembrane   70 -  86 (  58 -  92)
   INTEGRAL    Likelihood = -5.20   Transmembrane  105 - 121 ( 100 - 123)
   INTEGRAL    Likelihood = -4.25   Transmembrane  126 - 142 ( 123 - 144)
   INTEGRAL    Likelihood = -2.71   Transmembrane   18 -  34 (  18 -  34)

----- Final Results -----
                bacterial membrane --- Certainty=0.6243(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5031> which encodes the amino acid sequence <SEQ ID 5032>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
```

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.00   Transmembrane   69 -  85 (  62 -  93)
   INTEGRAL    Likelihood = -6.85   Transmembrane   16 -  32 (  11 -  37)
   INTEGRAL    Likelihood = -4.30   Transmembrane   99 - 115 (  96 - 121)
   INTEGRAL    Likelihood = -3.66   Transmembrane  126 - 142 ( 121 - 143)

----- Final Results -----
               bacterial membrane --- Certainty=0.6201(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 57/155 (36%), Positives = 96/155 (61%), Gaps = 5/155 (3%)

Query: 1   MVSYEKVRRSLRTATITIIVLNSLSLVFRLFTGISVQLAKTEI-NKGNTGNLPKEHIEAV 59
           M+SYEKVR++L+T+TI II+LN L +V  L     +   ++++ N+      L  E +  +
Sbjct: 1   MISYEKVRQALKTSTIAIIILNGLGVVLSLMGFAGIFYLQSQLKNEAPRAQLITEQLAQL 60

Query: 60  LSATTPFMLFVTALIVLVNIAIVIPCIKNLRAIKRNQTVNYLPYYLGFAITVGLVILGFL 119
            S+ TPFM+F++ L VL  IAI++PC +NL  +K+  TV+Y+PY LG  ++V  ++  F
Sbjct: 61  QSSMTPFMIFLSVLNVLAIIAIIVPCAQNLSKLKQGLTVSYIPYILGLILSVIGLVNQFT 120

Query: 120 TTKAPWAIAINIVFQAIFGLLYPHAYQKAQKLNER 154
           TT +     + ++  A++G     A+ KA+ LNE+
Sbjct: 121 TTMSMVGTILLILQAALYGF----AFYRAKTLNEK 151
```

SEQ ID 5030 (GBS227) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 119 (lane 5; MW 21.2kDa).

GBS227-His was purified as shown in Figure 227, lane 8-9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1630

A DNA sequence (GBSx1725) was identified in *S.agalactiae* <SEQ ID 5033> which encodes the amino acid sequence <SEQ ID 5034>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1224(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14706 GB:Z99118 similar to conjugation transfer protein
        [Bacillus subtilis]
 Identities = 328/754 (43%), Positives = 486/754 (63%), Gaps = 25/754 (3%)

Query: 2   EVFFTGTIERIIPRNASNFFKILLLEIRDTDSDFDDVKVIITGTMADVIEGEEYTFWGTL 61
           E +  GT+  +I+ N +N + +L +++ +T     +D  V +TG     + E E YTF+G +
Sbjct: 13  EPYLKGTVNTVIYTHNDTNLYTVLKVKVTETSEAIEDKAVSVTGYFPALQEEETYTFYGKI 72

Query: 62  TQHPKYGEQLQSVRYERAKPTSG-GLVKYFSSEQFKGIGKKTAQRIVELYGDNTIDKILE 120
             HPK+G Q Q+  +++  PT+  G+++Y SS+ F+GIGKKTA+ IV+  GD+ I+KIL
Sbjct: 73  VTHPKFGLQFQAEHFKKEIPTTKEGIIQYLSSDLFEGIGKKTAEIVKKLGDSAINKILA 132
```

```
Query: 121 SPEQLSTISGLSKINREAFIAKLKLNYGTEQVLAKLAEYGLSNRAAIQIFDHYKRESLEV 180
             L  +  LSK   +      L+ + G EQ++  L ++G   + +++I+  Y+ E+LE
Sbjct: 133 DASVLYDVPRLSKKKADTLAGALQRHQGLEQIMISLNQGFGPQLSMKIYQAYBSETLEK 192

Query: 181 INENPYQLVEDIQGIGFKIADQLAEQVGIESDSPKRFRAAIIHTLVESSMEQGDTYIEAR 240
             I ENPYQLV+D+ GIGF  AD+L  ++G+  + P+R +AAI++TL  + + +G TYIE
Sbjct: 193 IQENPYQLVKDVEGIGFGKADELGSRMGLSGNHPERVKAAILYTLETTCLSEGHTYIETE 252

Query: 241 TLLEKTITLLKRA------RQIKLDPS---IVAKELTNLIARDKVQHIGTKIFSNTLFFAS 292
             L+  T +LL ++       R  E+D +   I   E  +++ ED       + +  +LF+AE
Sbjct: 253 QLIIDTQSLLNQSARSGQRITPMDAANAIIALGRNKDIVIEDG------RCYPSLFYAE 306

Query: 293 EGIKKNLQRILNQP-LDKQLNHKDIDREIRDIQKSLNITHYDNIQEKAIREALLSKVPILT 351
             + + K ++ I +Q   + Q    +  ++++ +++ Y   Q++AI++AL S + +LT
Sbjct: 307 QNVAKRVKHIASQTEYENQFPSEFLLALGELEERMDVQYAPSQKEAIQKALSSPMLLLT 366

Query: 352 GGPGTGKTTVINGIIEAYSELHHIDLN----KND--IPIVLAAPTGRAARRMNELTGLPS 405
             GGPGTGKTTVI GI+E Y ELH + L+    K D   PIVLAAPTGRAA+RM+E TGLP+
Sbjct: 367 GGPGTGKTTVIRGIVELYGELHGVSLDPSAYKKDEAFPIVLAAPTGRAAKRMSESTGLPA 426

Query: 406 ATIHRHLGLNGDSDYQSLDDY-LDCSLIIIDEFSMVDTWLANQLFDALDSHTQVIIVGDS 464
             TIHR LG NG   +D  ++  L+IIDE SM+D WLAN LF A+  H Q+IIVGD
Sbjct: 427 VTIHRLLGWNGAEGFTHTEDQPIEGKLIIIDEASMLDINLANHLFKAIPDHIQIIIVGDE 486

Query: 465 DQLPSVGPGQVLADLLNINALPHVKLEKIFRQSEESTIVTLANQMRQGFLPEDFTAKKAD 524
             DQLPSVGPGQVL DLL    +P V+L  I+RQ+E S+IV LA+QM+ G LP + TA   D
Sbjct: 487 DQLPSVGPGQVLRDLLASQVIPTVRLTDIYRQAEGSSIVELAHQMKNGLLPNNLTAPTKD 546

Query: 525 RSYFEASANIIPNMISKIVQSALKSGIEAHRIQILAPMYRGQAGINNLNLIMQNLLNPLK 584
             RS+      + I  ++ K+V +ALK G  A +IQ+LAPMYRG+AGIN LN+++Q++LNP K
Sbjct: 547 RSFIRCGSSQIKSVVEKVVANALKKGYTAKDIQVLAPMYRGKAGINELNVMLQDILNPPK 606

Query: 585 D-NNQFTFNDINFRIGDKVLHLVNNTELNVPNGDIGYITDLIPAKYTESKQDEIYMTFDG 643
             +  F D+ +R GDK+L LVN  E NVFNGDIG IT +  AK   K+D   ++FDG
Sbjct: 607 EKRRELKFGDVYYRTGDKILQLVNQPENNVFNGDIGEITSIFYAKENTEKEDMAVVSFDG 666

Query: 644 QEVIYQRKEWLKITLAYAMSIHKSQGSEFQVVILPITRQSGRMLQRNLIYTATRSKSKLT 703
             E+ + +K++ + T AY  SIHKSQGSEF +V+LP+ +   RML+RNL+YTATR+K L
Sbjct: 667 NEMTFTKKDFNQFTHAYCCSIHKSQGSEFPIVVLPVKGYYRMLRRNLLYTAITRAKKFL 726

Query: 704 ILLGEIGAFDFAVKNEGAK-RNTYLIERFENKQE 736
             IL GE  A ++ VKN  A  R T L  R   + E
Sbjct: 727 ILCGEEEALEWGVKNNDATVRQTSLKNRLSVQVE 760
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5035> which encodes the amino acid sequence <SEQ ID 5036>. Analysis of this protein sequence reveals the following:

```
        Possible site: 47

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 232-234
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB14706 GB:Z99118 similar to conjugation transfer protein
            [Bacillus subtilis]
 Identities = 318/769 (41%), Positives = 473/769 (61%), Gaps = 29/769 (3%)

Query: 7   GTVDRIIFENQANFFKILLLAIEDTDSDIDDFEIITGTMADIIEGDDYTFWGELTQHPK 66
            GTV+ +I+ N  N + +L + + +T   I+D  + +TG    + E + YTF+G++  HPK
Sbjct: 18  GTVNTVIYHNDTNLYTVLKVKVTSTSEAIEDKAVSVTGYFPALQEETYTFYGKIVTHPK 77
```

```
Query: 67  YGQQLKLSRYQKIKPSSS-GLVNYFSSDHFKGIGKKTAEKIIALYGHNTIDHILEDPSKL 125
           +G Q +   ++K  P++  G++ Y SSD F+GIGKKTAE+I+   G + I+ IL D S L
Sbjct: 78  FGLQFQAEHFKKEIPTTKEGIIQYLSSDLFEGIGKKTAEEIVKKLGDSAINKILADASVL 137

Query: 126 ETISGLSKANRQAFVAKLKLNYGTEQLIAGLVELGLSNRFALQAFEKYKEEALDLVKENP 185
             +  LSK         L+ + G EQ++  L + G   + +++ ++ Y+ E L+ ++ENP
Sbjct: 138 YDVPRLSKKKADTLAGALQRHQGLEQIMISLNQFGFGPQLSMKIYQAYESETLEKIQENP 197

Query: 186 YQLVEDLQGFGFKMADALAENLGIESDSPKRFRAALLHCLLEESINRGDTYVQARQLLDF 245
           YQLV+D++G GF  AD L   +G+  + P+R +AA+L+ L    ++ G TY++  QL+
Sbjct: 198 YQLVKDVEGIGFGKADELGSRMGLSGNHPERVKAAILYTLETTCLSEGHTYIETEQLIID 257

Query: 246 AITLL-----EDARQVECDPAAVAEQLSE---LIIEGKIKNSDTKLFDASLYFAEEGIAN 297
             +LL     E  R  E D A     L E   ++IE      D + +  SL++AE+ +A
Sbjct: 258 TQSLLNQSAREGQRITEMDAANAIIALGENKDIVIE------DGRCYFPSLFYAEQNVAK 311

Query: 298 NISRLLD-TPLSQSFSHDTIQTTIQAVQKDFAITYDQVQQEAITKALTSKVFLLTGGPGT 356
            +  +   T     F        +  +++   + Y   Q+EAI KAL+S + LLTGGPGT
Sbjct: 312 RVKHIASQTEYENQFPESEFLLALGELEERMDVQYAPSQKEAIQKALSSPMLLLTGGPGT 371

Query: 357 GKTTVIRGILQAYANLHQIDLD----KKD--LPILLAAPTGRAARRMNELTGLPSATIHR 410
           GKTTVIRGI++ Y  LH + LD    KKD   PI+LAAPTGRAA+RM+E TGLP+ TIHR
Sbjct: 372 GKTTVIRGIVELYGELHGVSLDPSAYKKDEAFPIVLAAPTGRAAKRMSESTGLPAVTIHR 431

Query: 411 HLGLNGDNDYQAMEDY-LDCDLLIVDEFSMVDTWLANQLLGAINSTTQVIIVGDSDQLPS 469
            LG NG    +   ED  ++  LLI+DE SM+D WLAN L  AI    Q+IIVGD DQLPS
Sbjct: 432 LLGWNGAEGFTHTEDQPIEGKLLIIDEASMLDIWLANHLFKAIPDHIQIIIVGDEDQLPS 491

Query: 470 VGPGQVLSDLLKVNSLPQIALQKIFRQSQESTIVNLADQMRRGILAADFRDKKADRSYFE 529
           VGPGQVL DLL    +P + L  I+RQ++ S+IV LA QM+ G+L  +      DRS+
Sbjct: 492 VGPGQVLRDLLASQVIPTVRLTDIYRQAEGSSIVELAHQMKNGLLPNNLTAPTKDRSFIR 551

Query: 530 AQAAFIPDMIQKIVLSAIKSGIPAEEIQILAPMYKGQAGINHLNQLMQELLN-PLQGQTE 588
              + I ++++K+V +A+K G  A++IQ+LAPMY+G+AGIN LN ++Q++LN P + + E
Sbjct: 552 CGGSQIKEVVEKVVANALKKGYTAKDIQVLAPMYRGKAGINELNVMLQDILNPPKEKRRE 611

Query: 589 FLFNDTHFRKGDKVLHLVNDAQLNVFNGDIGYITDLIPAKYTESKQDELILDFDGSEVTY 648
             F D  +R GDK+L LVN  + NVFNGDIG IT +  AK    K+D  ++ FDG+E+T+
Sbjct: 612 LKFGDVVYRTGDKILQLVNQPENNVFNGDIGEITSIFYAKENTEKEDMAVVSFDGNEMTF 671

Query: 649 PRNEWLKLTLAYAMSIHKSQGSEFQVVILPITRQSGRLLQRNVIYTAITRSKSKLILLGE 708
            + ++ + T AY  SIHKSQGSEF +V+LP+ +   R+L+RN++YTAITR+K  LIL GE
Sbjct: 672 TKKDFNQFTHAYCCSIHKSQGSEFPIVVLPVVKGYYRMLRRNLLYTAITRAKKFLILCGE 731

Query: 709 YTAFEYAIK-HEGDKRQTYLIERFQEQSDLASSQPNQELKSKEQTSLFS 756
             A E+ +K ++   RQT L  R   Q +    + + EL++ ++   FS
Sbjct: 732 EEALEWGVKNNDATVRQTSLKNRLSVQVE----EMDAELEALQKELPFS 776
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 544/816 (66%), Positives = 665/816 (80%), Gaps = 10/816 (1%)

Query: 1   MEVFFTGTIERIIFENASNFFKILLLEIEDTDSDFDDVEVIITGTMADVIEGEEYTFWGT 60
           ME  FTGT++RIIFEN +NFFKILLL IEDTDSD DD E+IITGTMAD+IEG++YTFWG
Sbjct: 1   MEYVFTGTVDRIIFENQANFFKILLLAIEDTDSDIDDFEIIITGTMADIIEGDDYTFWGE 60

Query: 61  LTQHPKYGEQLQSVRYERAKPTSGGLVKYFSSEQFKGIGKKTAQRIVELYGDNTIDKILE 120
           LTQHPKYG+QL+  RY++ KP+S GLV YFSS+ FKGIGKKTA++I+ LYG NTID ILE
Sbjct: 61  LTQHPKYGQQLKLSRYQKIKPSSSGLVNYFSSDHFKGIGKKTAEKIIALYGHNTIDHILE 120

Query: 121 SPEQLSTISGLSKINREAFIAKLKLNYGTEQVLAKLAEYGLSNRAAIQIFDHYKEESLEV 180
            P +L TISGLSK NR+AF+AKLKLNYGTEQ++A L E GLSNR A+Q F+ YKEE+L++
Sbjct: 121 DPSKLETISGLSKANRQAFVAKLKLNYGTEQLIAGLVELGLSNRFALQAFEKYKEEALDL 180

Query: 181 INENPYQLVEDIQGIGFKIADQLAEQVGIESDSPKRFRAAIIHTLVESSMEQGDTYIEAR 240
           + ENPYQLVED+QG GFK+AD LAE +GIESDSPKRFRAA++H L+E S+ +GDTY++AR
Sbjct: 181 VKENPYQLVEDLQGFGFKMADALAENLGIESDSPKRFRAALLHCLLEESINRGDTYVQAR 240
```

```
Query: 241 TLLEKTTTLLEEARQIELDPSIVAKKLTNLIAEDKVCHIGTKIFSNTLFFAEEGIKKNLQ 300
           LL+  TTLLE+ARQ+E DP+ VA++L+ LI E K+++   TK+F  +L+FAEEGI  N+
Sbjct: 241 QLLDFAITLLEDARQVECDPAAVAEQLSELIIEGKIKNSDTKLFDASLYFAEEGIANNIS 300

Query: 301 RILNQPLDKQLNHKDIDREIRDIQKSLNTHYDNIQRKAIREALLSKVFLLTGPGTGKTT 360
           R+L+ PL +  +H  I   I+ +QK   I YD +Q++AI +AL SKVF+LTGPGTGKTT
Sbjct: 301 RLLDTPLSQSFSHDTIQTTIQAVQKDFAITYDQVQQSAITKALTSKVFLLTGPGTGKTT 360

Query: 361 VINGIIEAYSELHHIDLNKNDIPIVLAAPTGRAARRMNELTGLPSATIHRHLGLNGDSDY 420
           VI GI++AY+ LH IDL+K D+PI+LAAPTGRAARRMNELTGLPSATIHRHLGLNGD+DY
Sbjct: 361 VIRGILQAYANLHQIDLDKKDLPILLAAPTGRAARRMNELTGLPSATIHRHLGLNGDNDY 420

Query: 421 QSLDDYLDCSLIIIDEFSMVDTWLANQLFDALDSHTQVIIVGDSDQLPSVGPGQVLADLL 480
           Q+++DYLDC L+I+DEFSMVDTWLANQL  A++S TQVIIVGDSDQLPSVGPGQVL+DLL
Sbjct: 421 QAMRDYLDCDLLIVDEFSMVDTWLANQLIGAINSTTQVIIVGDSDQLPSVGPGQVLSDLL 480

Query: 481 NINALPHVKLEKIFRQSEESTIVTLANQMRQGFLPEDFTAKKADRSYFEASANIIPMMIS 540
            +N+LP + L+KIFRQS+ESTIV LA+QMR+G L  DF  KKADRSYFEA A  IP+MI
Sbjct: 481 KVNSLPQIALQKIFRQSQESTIVNLADQMRRGILAADFRDKKADRSYFEAQAAFIPDMIQ 540

Query: 541 KIVQSALKSGIEAHEIQILAPMYRGQAGINNLNLIMQNLLNPLKDNNQFTYNDINFRIGD 600
           KIV SA+KSGI A EIQILAPMY+GQAGIN+LN +MQ LLNPL+   +F FND +FR GD
Sbjct: 541 KIVLSAIKSGIPAEEIQILAPMYKGQAGINHLNQLMQELLNPLQGQTSFLFNDTHFRKGD 600

Query: 601 KVLHLVNDTRLNVFNGDIGYITDLIPAKYTRSKQDRIYMTFDGQRVIYQRKEWLKITLAY 660
           KVLHLVND +LNVFNGDIGYITDLIPAKYTESKQDE+ + FDG EV Y R EWLK+TLAY
Sbjct: 601 KVLHLVNDAQLNVFNGDIGYITDLIPAKYTESKQDELILDFDGSEVTYPRNEWLKLTLAY 660

Query: 661 AMSIHKSQGSEFQVVILPITRQSGRMLQRNLIYTAITRSKSKLLLLGEIGAFDFAVKNEG 720
           AMSIHKSQGSEFQVVILPITRQSGR+LQRN+IYTAITRSKSKLLLLGE  AF++A+K+EG
Sbjct: 661 AMSIHKSQGSEFQVVILPITRQSGRLLQRNVIYTAITRSKSKLLLLGSYTAFEYAIKHEG 720

Query: 721 AKRNTYLIERFENKQEIANSQKIEDSSIDQKI----------DNTIINTSIPKTATPIEQ 770
           KR TYLIERF+ + ++A+SQ  ++          ++          D++  ++S   + P E
Sbjct: 721 DKRQTYLIERFQEQSDLASSQPNQELKSKEQTSLFSNTATLEDDSQKSSSQSTNSNPTEN 780

Query: 771 TNLSKITYRLTSENYLTIDPMIGINQQDISAIFDSK 806
           +      +RLT ENY TID MIG+ + DI+  F  K
Sbjct: 781 SQSNDDFRLTPENYSTIDSMIGLTESDIALFFQKK 816
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1631

A DNA sequence (GBSx1726) was identified in *S.agalactiae* <SEQ ID 5037> which encodes the amino acid sequence <SEQ ID 5038>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.23   Transmembrane    9 -  25 (   7 -  29)

----- Final Results -----
               bacterial membrane --- Certainty=0.4291(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB69116 GB:U90721 signal peptidase I [Streptococcus pneumoniae]
 Identities = 120/201 (59%), Positives = 144/201 (70%), Gaps = 9/201 (4%)

Query: 2   KEFIKEWGVFILILSLFLLSRIFLWQFVKVDGHSMDPTLADKEQLVVLKQTKINRFDIVV 61
           K F+KEWG+F+LILSL  LSRIF W  V+V+GHSMDPTLAD E L V+K    I+RFDIVV
Sbjct: 5   KNFLKEWGLFLLILSLLALSRIFFWSNVRVEGHSMDPTLADGELLFVVKHLPIDRFDIVV 64
```

```
Query: 62  ANEEBGGQKKKIVKRVIGMPGDVIKYKNDTLTINNKKTEEPYLKEYTKLFKKDKLQEKYS 121
           A+EE+G    K IVKRVIGMPGD I+Y+ND L IN+K+T+EPYL +Y K FK DKLQ  YS
Sbjct: 65  AHEEDG--NKDIVKRVIGMPGDTIRYENDKLYINDKETDEPYLADYIKRFKDDKLQSTYS 122

Query: 122 -------YNPLFQDLAQSSTAFTTDSNGSSEFTTVVPKGHYYLVGDDRIVSKDSRAVGPF 174
                  F+ +AQ + AFT D N ++ F+  VP+G Y L+GDDR+VS DSR VG F
Sbjct: 123 GKGFEGNKGTFFRSIAQKAQAFTVDVNYINFSFTVPEGEKYLLLGDDRLVSSDSRHVGTF 182

Query: 175 KKSTIVGEVKPRFWPIRRFGT 195
           K   I GE KPRFWPI R GT
Sbjct: 183 KAKDITGEAKPRFWPITRIGT 203
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5039> which encodes the amino acid sequence <SEQ ID 5040>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.50   Transmembrane   35 -  51 (  35 -  51)

----- Final Results -----
                bacterial membrane --- Certainty=0.1999(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9157> which encodes the amino acid sequence <SEQ ID 9158>. Analysis of this protein sequence reveals the following:

```
      Possible site: 43
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                 bacterial outside --- Certainty= 0.300(Affirmative) < succ>
                bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 131/197 (66%), Positives = 152/197 (76%)

Query: 1   MKEFIKEWGVVFLILSLFLLSRIFLWQFVKVDGHSMDPTLADKEQLVVLKQTKINRFDIV 60
           MK+FIKEWG F L L LF LSR+FLWQ VKVDGHSMDPTLA  E+L+V  Q +I+RFDIV
Sbjct: 23  MKQFIKEWGPFTLFLILFGLSRLFLWQAVKVDGHSMDPTLAHGSRLIVPNQARIDRFDIV 82

Query: 61  VANEEGGQKKKIVKRVIGMPGDVIKYKNDTLTINNKKTEEPYLKEYTKLFKKDKLQEKY 120
           VA EE GQKK+IVKRVIG+PGD I Y +DTL IN KKT EPYL EY K FK DKLQ+ Y
Sbjct: 83  VAQEEENGQKKEIVKRVIGLPGDTISYNDDTLYINGKKTVEPYLAEYLKQFKNDKLQKTY 142

Query: 121 SYNPLFQDLAQSSTAFTTDSNGSSEFTTVVPKGHYYLVGDDRIVSKDSRAVGPFKKSTIV 180
           +YN LFQ LA++S AFT+S G + F   VPKG Y L+GDDRIVS+DSR VG FKK  ++
Sbjct: 143 AYNTLFQQLAETSDAFTTNSEGQTRFEMSVPKGRYLLLGDDRIVSRDSRHVGSFKKENLI 202

Query: 181 GEVKPRFWPIRRFGTIN 197
           GEVK RFWP+ +   N
Sbjct: 203 GEVSARFWPLNKMTVFN 219
```

SEQ ID 5038 (GBS268) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 4; MW 50.3kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 9; MW 25.3kDa) and in Figure 160 (lane 2-4; MW 25.3kDa).

GBS268-His was purified as shown in Figure 222, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1632

A DNA sequence (GBSx1727) was identified in *S.agalactiae* <SEQ ID 5041> which encodes the amino acid sequence <SEQ ID 5042>. This protein is predicted to be ribonuclease HIII (rnhB). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4728(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10119> which encodes amino acid sequence <SEQ ID 10120> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45437 GB:U93576 ribonuclease HII [Streptococcus pneumoniae]
 Identities = 176/282 (62%), Positives = 219/282 (77%), Gaps = 13/282 (4%)

Query: 16  EKIRTDLAQHHISNNNPYVVPSAKISGATVLLYTSGKLVFQGSNASHIAQKYGF--IEQK 73
           E  +T LA     + NPY+ +  K+  ATV +YTSGK++ QG  A   A  +G+  +EQ
Sbjct: 18  EHYQTSLAP----SKNPYIRYFLKLPQATVSIYTSGKILLQGEGAEKVASFFGYQAVEQ- 72

Query: 74  ESCSSSSQDIPIIGTDEVGNGSYFGGLAVVASFVTPKDHAYLKKLGVGDSKTLTDQKIKQ 133
               +  Q++P+IGTDEVGNGSYFGGLAVVA+FVTP  H +L+KLGVGDSKTLTDQKI+Q
Sbjct: 73  ----TSQQNLPLIGTDEVGNGSYFGGLAVVAAFVTPDQHDFLRKLGVGDSKTLTDQKIRQ 128

Query: 134 IAPLLEKAIPHKALLLSPQKYNQVVSPNNKHNAVSVKVALHNQAIFLLLQGFEPEKIVI 193
           IAP+L++ I H+ALLLSP KYN+V+    +++NAVSVKVALHNQAI+LLLQ G +PEKIVI
Sbjct: 129 IAPILKEKIQHQALLLSPSKYNEVIG--DRYNAVSVKVALHNQAIYLLLQKGVQPEKIVI 186

Query: 194 DAFTSSKNYQNYLKNEKNQFKQTITLEEKAENKYLAVAVSSIIARNLFLENLNKLSDVVG 253
           DAFTS+KNY  YL  E N+F   I+LEEKAE KYLAVAVSS+IAR+LFLENL  L  ++G
Sbjct: 187 DAFTSAKNYDKYLAQETNRFSNPISLEEKAEGKYLAVAVSSVIARDLFLENLENLGRELG 246

Query: 254 YKLPSGAGHQSDKVASQLLKAYGISSLEHCAKLHFANTKKAQ 295
           Y+LPSGAG  SDKVASQ+L+AYG+  L  CAKLHF NT+KA+
Sbjct: 247 YQLPSGAGTASDKVASQILQAYGMQGLNFCAKLHFKNTRKAK 288
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5043> which encodes the amino acid sequence <SEQ ID 5044>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2148(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 194/298 (65%), Positives = 240/298 (80%), Gaps = 2/298 (0%)

Query: 3   MNTIVMQADKKLQEKIRTDLAQHHISNNNPYVVPSAKISGATVLLYTSGKLVFQGSNASH 62
           MNT+V++ D  L + ++  LA + IS+ N YV P+AK +G TVLLY SGKLV QG+ A+
Sbjct: 1   MNTLVLKIDALLSKHLKKQLAPYTISSQNTYVAPAAKKNGVTVLLYKSGKLVLQGNGANA 60
```

```
Query: 63  IAQKYGFIEQKE--SCSSESQDIPIIGTDEVGNGSYPGGLAVVASFVTPKDHAYLKKLGV 120
           +AQ+       K      S+ SQDIPIIG+DEVGNGSYPGG+AVVASFV PKDH++LKKLGV
Sbjct: 61  LAQSLNLPVAKTVPEASNNSQDIPIIGSDEVGNGSYPGGIAVVASFVDPKDHSFLKKLGV 120

Query: 121 GDSKTLTDQKIKQIAPLLEKAIPHKALLLSPQKYNQVVSPNNKHNAVSVKVALHNQAIFL 180
            DSK L+D+ I+QIAPLLEK IPH++LLLSP+KYN++V  +  +NA+S+KVALHNQAIFL
Sbjct: 121 DDSKKLSDKTIQQIAPLLEKQIPHQSLLLSPKKYNELVGKSKPVNAISIKVALHNQAIFL 180

Query: 181 LLQDGFEPEKKIVIDAFTSSKNYQNYLKNEKNQFKQTITLERKAENKYLAVAVSSIIARNL 240
           LLQ G +P++IVIDAFTS  NY+ +LK EKN F   +T +EKAE+ YLAVAVSSIIARNL
Sbjct: 181 LLQKGIQPKQIVIDAFTSQSNYSKHLKKEKNHFPNPLTFQEKAESHYLAVAVSSIIARNL 240

Query: 241 FLENLNKLSDDVGYKLPSGAGHQSDKVASQLLKAYGISSLEHCAKLHFANTKKQALL 298
           FL+NL++L  D+GY+LPSGAG  SDKVASQLL AYG+SSLE+ AKLHFANT KQALL
Sbjct: 241 FLDNLDQLSQDLGYQLPSGAGSASDKVASQLLAAYGMSSLEYSAKLHFANTHKQALL 298
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1633

A DNA sequence (GBSx1728) was identified in *S.agalactiae* <SEQ ID 5045> which encodes the amino acid sequence <SEQ ID 5046>. This protein is predicted to be heat shock protein 70. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3874(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5047> which encodes the amino acid sequence <SEQ ID 5048>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3442(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 65/92 (70%), Positives = 76/92 (81%)

Query: 11  NRYKFVFGDKPLILTTDKDNLFMEEIERVATEKYEAIEKELPNADNSTTAILMAINALSV 70
           NRYKF FG+K LILTTDKDNLFMEE+ERVA EKY+A+K  LP AD+ETIAILMAIN LS
Sbjct: 5   NRYKFTFGEKTLILTTDKDNLFMEEVERVAKEKYQALKNHLPRADDETIAILMAINTLST 64

Query: 71  QLSREIDIEKMEDELNKLRSKTISDIKKKVSE 102
           QLSREI  EKME E+  LR KT+  ++EK ++
Sbjct: 65  QLSREIAIEKMEAEILDLRQKTLVGLQKKANQ 96
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1634

A DNA sequence (GBSx1729) was identified in *S.agalactiae* <SEQ ID 5049> which encodes the amino acid sequence <SEQ ID 5050>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.99   Transmembrane  124 - 140 ( 114 - 148)
   INTEGRAL    Likelihood = -5.84   Transmembrane   22 -  38 (  21 -  40)
   INTEGRAL    Likelihood = -4.88   Transmembrane    2 -  18 (   1 -  20)
   INTEGRAL    Likelihood = -1.97   Transmembrane   84 - 100 (  84 - 100)

----- Final Results -----
                bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06827 GB:AP001517 unknown conserved protein in B. subtilis
        [Bacillus halodurans]
 Identities = 59/182 (32%), Positives = 98/182 (53%), Gaps = 14/182 (7%)

Query: 1   MLSLLLLIIVIWHFYIGYSRGIFLQVFYVLMSMVSLMIASQFYQELASQITLWVPYS--N 58
            MLS++LL I++  F+IG  RG+ LQ+ ++L  + +  +A ++Y  +A+ I LW+PY   +
Sbjct: 1   MLSVILLFILLCSFFIGKRRGLILQLVHLLGFVAAPFVAYKYYAPVATYIRLWIPYPQPS 60

Query: 59  PVQGVEVYFFKDISKFQLSHVYYAGVAFVFIY----SLSYLVGRLLGVLLHLAPVEHFDS 114
            P   V +      I  F    +VYY+G+AF  ++       L ++VG +L  L HL  +
Sbjct: 61  PDSFVTML----IEAFNFENVYYSGIAFALLFIGTKILLHIVGSMLDFLTHLPILRSV-- 114

Query: 115 LQNNIISGFLAVLVCLLFMSMCLTILATVPMSFVQEKLWNSLSVFRFLINDLDFFSQFLVR 174
             N  + G L  +    L M + L + A +P+  VQ  L  SL  +P++N  PF S+F+
Sbjct: 115 --NGWLGGILGFVEVVLIMPVLLYVGALLPIETVQTHLNQSLVAQFIMNHTPFLSEFIRN 172

Query: 175 TW 176
             W
Sbjct: 173 LW 174
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5051> which encodes the amino acid sequence <SEQ ID 5052>. Analysis of this protein sequence reveals the following:

```
      Possible site: 59
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.17   Transmembrane  124 - 140 ( 117 - 148)
   INTEGRAL    Likelihood = -4.73   Transmembrane   84 - 100 (  78 - 105)
   INTEGRAL    Likelihood = -0.00   Transmembrane  156 - 172 ( 156 - 172)

----- Final Results -----
                bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06827 GB:AP001517 unknown conserved protein in B. subtilis
        [Bacillus halodurans]
 Identities = 57/177 (32%), Positives = 98/177 (55%), Gaps = 2/177 (1%)

Query: 1   MLSLLIVLILTWNFYIGYSRGILQSFYVLGALLSLLVANRFYIGLAHKLTLWIPYSNPV 60
            MLS++++ IL  +F+IG  RG+ILQ  ++LG + +  VA ++Y  +A  + LWIPY
Sbjct: 1   MLSVILLFILLCSFFIGKRRGLILQLVHLLGFVAAPFVAYKYYAPVATYIRLWIPYPQPS 60

Query: 61  EGTSVFFFKSVDIFVLDKVYYAGLAFPIIFLLGYALSRFLGIFVHFLLLNYFDNQWTKCL 120
             + V      ++ F  + VYY+G+AF ++F+       L  +G  + FL          L
Sbjct: 61  PDSFVTML--IEAFNFENVYYSGIAFALLFIGTKILLHIVGSMLDFLTHLPILRSVNGWL 118
```

```
Query: 121 SGGLAFLVSLLFLNMLLSIFATVPMPFLQHYLHSSFLARLVIEHLPPLTIIIQKLWI 177
           G L F+   L + +LL + A +P+  +Q +L+ S +A+ ++ H P L+  I+ LWI
Sbjct: 119 GGILGFVEVYLIMFVLLYVGALLPISTVQTHLNQSLVAQFIMNHTPFLSSFIHNLWI 175
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 87/176 (49%), Positives = 123/176 (69%)

Query: 1   MLSLLLLIIVIWHFYIGVSRGIPLQVFYVLMSMVSLMIASQFYQELASQITLWVPYSNPV 60
           MLSLL+++I+ W+FYIGYSRGI LQ FYVL +++SL++A++FY  LA ++TLW+PYSNPV
Sbjct: 1   MLSLLIVLILVWNFYIGYSRGIILQSFYVLGALLSLLVANRFYIGLAHKLTLWIPYSNPV 60

Query: 61  QGVEVYFFKDISKFQLSHVYYAGVAFVFIYSLSYLVGRLLGVLLHLAPVSHHPDSLQNNII 120
           +G  V+FFK +  F L  VYYAG+AF  I+ L Y + R LG+ +H   + +FD+     +
Sbjct: 61  EGTSVFFFKSVDIFVLDKVYYAGIAFFIILGYALSRFLGIFVHFLLLNVFDNQWTKCL 120

Query: 121 SGFLAVLVCLLFMSMCLTILATVPMSFVQEKLWNSLFVRFLINDLPFPSQFLVRTW 176
           SG LA LV LLF++M L+I ATVPM F+Q  L +S   R +I  LP  +  + + W
Sbjct: 121 SGGLAFLVSLLFLNMLLSIFATVPMPFLQHYLHSSFLARLVIEHLPPLTIIIQKLW 176
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1635

A DNA sequence (GBSx1730) was identified in *S.agalactiae* <SEQ ID 5053> which encodes the amino acid sequence <SEQ ID 5054>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4176(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10117> which encodes amino acid sequence <SEQ ID 10118> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14818 GB:Z99118 similar to DNA mismatch repair protein
           [Bacillus subtilis]
 Identities = 320/790 (40%), Positives = 466/790 (58%), Gaps = 18/790 (2%)

Query: 10  MNNKILEQLEPNKVKELLLPYLKTRQQQERLSRLEPMTEAPKIEKSFNEISDMEQIFVEH 69
           M  K+L  LEF+KVKE ++ +  +   +E L EL+P      +I+K  +E+ +   I
Sbjct: 1   MQQKVLSALEPHKVKEQVIGHAASSLGKEMLLELKPSASIDEIKKQLDEVDEASDIIRLR 60

Query: 70  HSFGIVSLSSISESLKRLELSADLNIQELLAIKKVLQSSSDMIHFYSDL--DNVSPQSLD 127
                L  I  +L+R E+ + L+  E   I  +L +   M HF + +  D V     +
Sbjct: 61  GQAFPGGLVDIRGALRRAEIGSVLSPSEFTEISGLLYAVKQMKHFITQMAEDGVDIPLIH 120

Query: 128 RLFENLEQFNLQGSFQA-INDGGFLSHPASPELERIRRQLTNSERRVRQILQLMLKEKA 186
           + E L   +L+    + I+D G +   AS  L  IR QL   E RVR  L+ ML+  +
Sbjct: 121 QHAEQLITLSDLERDINSCIDDHGEVLDHASTTLRGIRTQLRTLESRVRDRLESMLRSSS 180

Query: 187 --ELLSENLIASRSGRSGVLPVKNTYRNRISGVVHDISSSGSTVYIEPRAVVTLNEEITQL 244
             ++LS+ ++  R+ R V+PVK  YR+   G+VHD SSSG+T++IEP+A+V +N  + Q
Sbjct: 181 ASKWLSDTIVTIRNDRFVIPVKQEYRSSYGGIVHDTSSSGMTLFIEPQAIVEMNNSLQQA 240

Query: 245 RADERHEESRILHAFSDLLRPHVATIRNNAWILGHLDFVRAKYLFMSDNKATIPEISNDS 304
           + E+ E  RIL   ++   + + +L  LDF+ AK  +   KAT P +++
Sbjct: 241 KVKEKQRIERILRVLTEKTAEYTEELFLDLQVLQTLDFIFAKARYAKAVKATKPIMNDTG 300
```

```
Query: 305 TLALINVRHPLL--SNPVANDLHFDQDLTAIVTTGPNTGGKTIMLKTLGLAQLMGQSGLP 362
            + L   RHPLL      VAND+   +D + IVTTGPNTGGKT+ LKTLGL  LM QSGL
Sbjct: 301 FIRLKKAPHPLLPPDQVVANDIELGRDFSTIVTTGPNTGGKTVTLKTLGLLTLMAQSGLH 360

Query: 363 VLADKGSKIAVFNNIFADIGDEQSIEQSLSTFSSHMTHIVSILNKADHNSLVLFDELGAG 422
            + AD+GS+ AVF ++FADIGDEQSIEQSLSTFSSHM +IV IL + + NSLVLFDELGAG
Sbjct: 361 IPADRGSEAAVFEHVFADIGDEQSIEQSLSTFSSHMVNIVGILEQVNENSLVLFDELGAG 420

Query: 423 TDPQEGAALAMAILEHLRLSNIKTMATTHYPELKAYGIETNFVENASMEFDAETLSPTYR 482
            TDPQEGA+LAM+IL+ +  +N + +ATTHYPELKAYG     V NAS+EFD ETLSPTY+
Sbjct: 421 TDPQEGAALAMSILDUVHRTNARVLATTHYPELKAYGYNREGVMNASVEFDIETLSPTYK 480

Query: 483 FMQGVPGRSNAFEIASRLGLAPFIVKQAK-QMTDSDSDVNRIIEQLEAQTLETRRRLDHI 541
             + GVPGRSNAFEI+ RLGL   I+ QAK +MT   ++V+ +I  LE         L
Sbjct: 481 LLIGVPGRSNAFEISKRLGLPDHIIGQAKSEMTATEHNEVDTMIASLEQSKKRAEEELSET 540

Query: 542 KEVEQENLKFNRAVKKLYNEPSHERDKELEKIYQSAQEIVDMALNESDTILKKL----ND 597
            + + +E  K ++ +++    E + ++DK LE+  Q+A R V  A+ E++ I+ +L     +
Sbjct: 541 RSIKEAEKLHKELCQQIIELNSKKDRMLEEAEQQAAEKVKAAMKEAEDIIHELRTIKEE 600

Query: 598 KSQLKVHEIIDAKAQIKKLAPQVDLSVMKVLNKAKKIKAARAPRIGDDIIVISYGQRGTL 657
              K HE+I+AK +++    P  + SK     K +K    R  + GD+ V ++QQ+GTL
Sbjct: 601 HKSFKDHELINAKKRLEGAMPAPEKSKKPEKPKTQK----RDFKPGDEVKVLTPQQKGTL 656

Query: 658 TSQLKDGRWEAQVGIIKMTLTQDEFTLVRVQEDQKVKSKQINVVKADSSGPRARLDLRG 717
            +     W  Q+GI+KM + + +   ++  E K K K I  VK  D      LDLRG
Sbjct: 657 LSKTGGMWNVQIGILKMKVKEKDLEPISAPEPK-KEKMITAVKGKDYH-VSLELDLRG 714

Query: 718 KRYEEAMQELDNFIDQALLNNMQQVDIIHGIGTGVIREGVTKYLRRNKHVKHFAYAPQNA 777
            +EYE A+  ++ ++D A+L    +V IIHG GTG +R+GV   L+ ++ VK   +
Sbjct: 715 ERYENALSRVSKYLDDAVLAGYPRVSIIHGKGTGALRKGVQDLLKNHRSVKSSRPGEAGE 774

Query: 778 GSSGATIVTL 787
            GSSG T+V L
Sbjct: 775 GSSGVTVVEL 784
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5055> which encodes the amino acid sequence <SEQ ID 5056>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3843(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 775/787 (98%), Positives = 781/787 (98%)

Query: 2   INLGIMKSMNNKILEQLEFNKVKELILPYLKTEQSQKELSELEPMTSAPKIKKSPNEISD 61
            I LGIMKSMNNKILEQLEFNKVKKL+LPYLKTEQSQKKL ELEPMTSAPKIKKSFNSISD
Sbjct: 32  IILGIMKSMNNKILEQLEFNKVKELLLPYLKTEQSQKELLELEPMTEAPKIKKSFNEISD 91

Query: 62  MEQIFVEHHSPGIVSLSSISESLKRLELSADLNIQELLAIKKVLQSSSDMIHFYSDLDNV 121
            MEQIFVEHHSPGIVSLSSISESLKRLELS DLNIQELLAIKKVLQSSSDMIHFYSDLDNV
Sbjct: 92  MEQIFVEHHSPGIVSLSSISESLKRLELSTDLNIQELLAIKKVLQSSSDMIHFYSDLDNV 151

Query: 122 SFQSLDRLFENLEQFPNLQGSFQAINDGGFLEHFASPELERIRRQLTNSERRVRQIIQDM 181
            SFQSLDRLFENLEQFPNLQGSFQAINDGGFLEHFASPELERIRRQLTNSERRVRQIIQDM
Sbjct: 152 SFQSLDRLFENLEQFPNLQGSFQAINDGGFLEHFASPELERIRRQLTNSERRVRQIIQDM 211

Query: 182 LKEKAELLSENLIASRSGRSVLPVKNTYRNRISGVVHDISSSGSTVYIEPRAVVTLNERI 241
            LKEKAELLSENLIASRSGRSVLPVKNTYRNRISGVVHDISSSGSTVYIEPRAVVTLNERI
Sbjct: 212 LKEKAELLSENLIASRSGRSVLPVKNTYRNRISGVVHDISSSGSTVYIEPRAVVTLNERI 271
```

```
Query: 242 TQLRADERHEESRILHAPSDLLRPHVATIRNNAWILGHLDFVRAKYLPMSDNKATIPRIS 301
           TQLRADERHEE RILHAPSDLLRPHVATIRNNAWILGHLDFVRAKYLPMSDNKATIP+IS
Sbjct: 272 TQLRADERHEEGRILHAPSDLLRPHVATIRNNAWILGHLDFVRAKYLPMSDNKATIPKIS 331

Query: 302 NDSTLALINVRHPLLSNPVANDLHPDQDLTAIVITGPNTGGKTIMLKTLGLAQLMGQSGL 361
           NDSTLALINVRHPLLSNPVANDLHFD DLTAIVITGPNTGGKTIMLKTLGLAQLMGQSGL
Sbjct: 332 NDSTLALINVRHPLLSNPVANDLHFDHDLTAIVITGPNTGGKTIMLKTLGLAQLMGQSGL 391

Query: 362 PVLADKGSKIAVFNNIFADIGDEQSIEQSLSTFSSHMTHIVSILNEADHNSLVLFDELGA 421
           PVLADKGSKIAVFNNIFADIGDEQSIEQSLSTFSSHMTHIVSILNEADHNSLVLFDELGA
Sbjct: 392 PVLADKGSKIAVFNNIFADIGDEQSIEQSLSTFSSHMTHIVSILNEADHNSLVLFDELGA 451

Query: 422 GTDPQEGASLAMAILEHLRLSNIKTMATTHYPELKAYGIETNFVENASMEFDAETLSPTY 481
           GTDPQEGASLAMAILEHLRLS+IKTMATTHYPELKAYGIETNFVENASMEFDAETLSPTY
Sbjct: 452 GTDPQEGASLAMAILEHLRLSHIKTMATTHYPELKAYGIETNFVENASMEFDAETLSPTY 511

Query: 482 RFWQGVPGRSNAFEIASRLGLAPFIVKQAKQMTDSDSDVNRIIEQLEAQTLETRRRLDHI 541
           RFWQGVPGRSNAFEIASRLGLAPFIVKQAKQMTDSDSDVNRIIEQLEAQTLETRRRLDHI
Sbjct: 512 RFWQGVPGRSNAFEIASRLGLAPFIVKQAKQMTDSDSDVNRIIEQLEAQTLETRRRLDHI 571

Query: 542 KEVEQENLKFNRAVKKLVNEPSHERDKELEKIYQEAQEIVDMALNESDTILKKLNDKSQL 601
           KEVEQENLKFNRAVKKLVNEPSHERDKELEKIYQEAQEIVDMALNESDTILKKLNDKSQL
Sbjct: 572 KEVEQENLKFNRAVKKLVNEPSHERDKELEKIYQEAQEIVDMALNESDTILKKLNDKSQL 631

Query: 602 KPHEIIDAKAQIKKLAPQVDLSKNKVLNKAKKIKAARAPRIGDDIIVTSYGQRGTLTSQL 661
           KPHEIIDAKAQIKKLAPQVDLSKNKVLNKAKKIKAARAPRIGDDIIVTSYGQRGTLTSQL
Sbjct: 632 KPHEIIDAKAQIKKLAPQVDLSKNKVLNKAKKIKAARAPRIGDDIIVTSYGQRGTLTSQL 691

Query: 662 KDGRWEAQVGIIKMTLTQDEFTLVRVQEEQKVKSKQINVVKKADSGGPRARLDLRGKRYE 721
           KDGRWEAQVGIIKMTLTQDEF+LVRVQEEQKVK+KQINVVKKAD SGPRARLDLRGKRYE
Sbjct: 692 KDGRWEAQVGIIKMTLTQDEFSLVRVQEEQKVKNKQINVVKKADSGGPRARLDLRGKRYE 751

Query: 722 EAMQELDNFIDQALLNNMGQVDIIHGIGTGVIREGVTKYLRRNKHVKHFAYAPQNAGSG 781
           EAMQELD+FIDQALLNNMGQVDIIHGIGTGVIREGVTKYLRRNKHVKHFAYAPQNAGSG
Sbjct: 752 EAMQELDHFIDQALLNNMGQVDIIHGIGTGVIREGVTKYLRRNKHVKHFAYAPQNAGSG 811

Query: 782 ATIVTLG 788
           ATIVTLG
Sbjct: 812 ATIVTLG 818
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1636

A DNA sequence (GBSx1731) was identified in *S.agalactiae* <SEQ ID 5057> which encodes the amino acid sequence <SEQ ID 5058>. This protein is predicted to be thioredoxin (trxA). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2721(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10115> which encodes amino acid sequence <SEQ ID 10116> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB40815 GB:AJ133006 thioredoxin [Listeria monocytogenes] (ver
```

```
          2)
 Identities = 64/100 (64%), Positives = 78/100 (78%), Gaps = 1/100 (1%)

Query: 15  MALEVTDATFVEETKEGLVLIDFWATWCGPCRMQAPILEQLSQKIDEDELKILKMDVDEN 74
           M  E+TDATF +ET EGLVL DFWATWCGPCRM AP+LE++ +E  E  LKI+KMDVDEN
Sbjct: 1   MVKRITDATFEQETSEGLVLTDFWATWCGPCRMVAPVLEEIQSSRGE-ALKIVKMDVDRN 59

Query: 75  PETARQFGIMSIPTLMFKKDGEVVKQVAGVHTKDQLKAII 114
           PET   PG+MSIPTL+ KKDGEVV+ + G   K++L  +I
Sbjct: 60  PETPGSPGVMSIPTLLIKKDGEVVETIIGYRPKHELDEVI 99
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5059> which encodes the amino acid sequence <SEQ ID 5060>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2721(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1637

A DNA sequence (GBSx1732) was identified in *S.agalactiae* <SEQ ID 5061> which encodes the amino acid sequence <SEQ ID 5062>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.54   Transmembrane  170 - 186 ( 167 - 191)
   INTEGRAL    Likelihood = -5.52   Transmembrane   87 - 103 (  86 - 107)
   INTEGRAL    Likelihood = -4.62   Transmembrane  105 - 121 ( 104 - 126)

----- Final Results -----
              bacterial membrane --- Certainty=0.4015(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA60798 GB:X67369 ORF3 [Clostridium perfringens]
 Identities = 27/67 (40%), Positives = 52/67 (77%)

Query: 1  MEIGQQIIRYRKQQALSQEELAEKVYVSRQEISNWENDKTYPDIHSLLLLSQIFQVSLDQ 60
          M++ +++    RK++ LSQE+LAEK+ +SRQ++S WE+ ++ PD++ L++LS+++ V++D
Sbjct: 1  MKLAEKLQLMRKRKGLSQEDLARKLGISRQAVSKWESGQSVPDLNKLIILSELYNVTIDY 60

Query: 61 LIKGDIE 67
          L+K  E
Sbjct: 61 LVKETYE 67
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1739> which encodes the amino acid sequence <SEQ ID 1740>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.86   Transmembrane  173 - 189 ( 169 - 194)
   INTEGRAL    Likelihood = -5.52   Transmembrane   90 - 106 (  89 - 110)
   INTEGRAL    Likelihood = -4.62   Transmembrane  108 - 124 ( 107 - 129)

----- Final Results -----
```

```
                bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 187/195 (95%), Positives = 191/195 (97%)

Query: 1   MEIGQQIIRYRKQQALSQEELAEKVYVSRQSISNWENDKTYPDIHSLLLLSQIFQVSLDQ 60
           MEIGQQIIRYRKQQALSQE+LAEKVYVSRQSISNWENDKTYPDIHSLLLLSQIFQVSLDQ
Sbjct: 4   MEIGQQIIRYRKQQALSQEKLAEKVYVSRQSISNWENDKTYPDIHSLLLLSQIFQVSLDQ 63

Query: 61  LIKGDIEKMKYTITQVDKKNFERDTKVMVTLMILLMISSYPLVYPLEWLGLGIFVLLSII 120
           LIKGDIEKMKYTITQVDKKNF+RDTKVMVTLMILLMISSYPLVYPLEWLGLGIFVLLSII
Sbjct: 64  LIKGDIEKMKYTITQVDKKNFRRDTKVMVTLMILLMISSYPLVYPLEWLGLGIFVLLSII 123

Query: 121 TMTYANRVERPKKKYDVQTYKEILAVSSGKLLDEIEKRERRAKLPYQKPLIVIVFFLITV 180
           TMTYANRVERPKKKYDVQ YKEILAVS+GKLLDEIEKRERRA LPYQKPLIVIVFFLITV
Sbjct: 124 TMTYANRVERPKKKYDVQFYKEILAVSNGKLLDEIEKRERRATLPYQKPLIVIVFFLITV 183

Query: 181 ATFFASRFIFTWLFH 195
           A  FASRF+FTWLFH
Sbjct: 184 AFAFASRFMFTWLFH 198
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1638

A DNA sequence (GBSx1733) was identified in *S.agalactiae* <SEQ ID 5063> which encodes the amino acid sequence <SEQ ID 5064>. This protein is predicted to be adenine glycosylase (mutY). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2385(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9425> which encodes amino acid sequence <SEQ ID 9426> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04650 GB:AP001510 adenine glycosylase [Bacillus halodurans]
 Identities = 130/331 (39%), Positives = 190/331 (57%), Gaps = 15/331 (4%)

Query: 1   MLQQTQVNTVIPYYKRFLEWFPQIKDLADAPEEQLLKAWEGLGYYSRVRNMQKAAQQVMV 60
           MLQQT+V+TVIPYY+ F+  FP ++ LA A E+Q+LKAWEGLGYYSR RN+Q A ++V+
Sbjct: 45  MLQQTRVITVIPYYQAFMRQFPTLETLAYAEKDQVLKAWEGLGYYSRARNLQSAVRSVVE 104

Query: 61  DPGGIFPHTYDDIASLKGIGPYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP 120
            +GG  P T  +I+ LKG+GPYTAGAI SI+++ PEPAVDGNVMRV++R+  +  DI
Sbjct: 105 SYGGEVDSTRKEISKLKGVGPYTAGAILSIAYDQPEPAVDGNVMRVLSRVLYIERDIAKV 164

Query: 121 KNRKIFQAIMEILIDPDRPGDFNQALMDLGTDISAKTPRPDESPIRFFNAAYLNGTYSK 180
           K R +F++++  LI  + P  FNQ LM+LG  + +  +P      P+R    A+  G  +
Sbjct: 165 KTRTLFESLLYDLISKENPGFFNQGLMELGALVCTPTSPGCLLCPVRDHCRAFAAGVQEQ 224

Query: 181 YPIKNTKKKPKPMRIQAFVIRNQNGQYLLEKNTKGRLLGGFWSFPIIETSPLSQQLLDFD 240
            PIK  KKKPK  ++ A VIRN+ GQ L+E+  +  LL   W FP +E          L
Sbjct: 225 LPIKAKKKKPKAKQLIAAVIRNEKGQVLIERRPEKGLLAKLWQFPNVE---------LES 275
```

```
Query: 241 DNQSNPIIWQTQNETFQREYQLKPQWTDNHFPNIKHTFSHQKWTIELIEGVVKAT-DLPN 299
             +  ++    +E F  +  +       +   ++H FSH  W I + E  VK    L +
Sbjct: 276 TKNAQQVLGDYIHERPHLDAAV-----GEVVQTVEHVFSHLIWNIRVYEATVKGVPSLND 330

Query: 300 APHLKWVAIEDFSLYPFATPQKKMLETYLKQ 330
              WV      Y F    +K+++  L++
Sbjct: 331 KYEADWVDDRTIENYAFPVSHQKIIQGNLRK 361
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5065> which encodes the amino acid sequence <SEQ ID 5066>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3579(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 330/333 (99%), Positives = 331/333 (99%)

Query: 1   MLQQTQVNTVIPYYKRFLEWFPQIKDLADAPEEQLLKAWEGLGYYSRVRNMQKAAQQVMV 60
            MLQQTQVNTVIPYYKRFLEWFPQIKDLADAPEEQLLKAWEGLGYYSRVRNMQKAAQQVMV
Sbjct: 52  MLQQTQVNTVIPYYKRFLEWFPQIKDLADAPEEQLLKAWEGLGYYSRVRNMQKAAQQVMV 111

Query: 61  DFGGIFPHTYDDIASLKGIGPYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP 120
            DFGGIFPHTYDDIASLKGIGFYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP
Sbjct: 112 DFGGIFPHTYDDIASLKGIGFYTAGAIASISFNLPEPAVDGNVMRVMARLFEVNYDIGDP 171

Query: 121 KNRKIFQAIMELIDPDRPGDFNQALMDLGTDIESAKTPRPDESPIRFFNAAYLNGTYSK 180
            KNRKIFQAIMELIDPDRPGDFNQALMDLGTDIESAKTPRPDESPIRFFNAAYLNGTY K
Sbjct: 172 KNRKIFQAIMELIDPDRPGDFNQALMDLGTDIESAKTPRPDESPIRFFNAAYLNGTYGK 231

Query: 181 YPIKNTKKKPKPMRIQAFVIRNQNGQYLLEKNTKGRLLGGFWSFPIIETSPLSQQLDLFD 240
            YPIKN KKKPKPMRIQAFVIRNQNGQYLLEKNTKGRLLGGFWSFPIIETSPLSQQLDLFD
Sbjct: 232 YPIKNPKKKPKPMRIQAFVIRNQNGQYLLEKNTKGRLLGGFWSFPIIETSPLSQQLDLFD 291

Query: 241 DNQSNPIIWQTQNETFQREYQLKPQWTDNHFPNIKHTFSHQKWTIELIEGVVKATDLPNA 300
            DNQSNPIIWQTQNETF+REYQLKPQWTDNHFPNIKHTFSHQKWTIELIEGVVKATDLPNA
Sbjct: 292 DNQSNPIIWQTQNETFEREYQLKPQWTDNHFPNIKHTFSHQKWTIELIEGVVKATDLPNA 351

Query: 301 PHLKWVAIEDFSLYPFATPQKKMLETYLKQKNA 333
            PHLKWVAIEDFSLYPFATPQKKMLETYLKQKNA
Sbjct: 352 PHLKWVAIEDFSLYPFATPQKKMLETYLKQKNA 384
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1639

A DNA sequence (GBSx1734) was identified in *S.agalactiae* <SEQ ID 5067> which encodes the amino acid sequence <SEQ ID 5068>. This protein is predicted to be maltose/maltodextrin transport system (malG). Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.30   Transmembrane   14 -  30 (   5 -  35)
   INTEGRAL    Likelihood = -6.95   Transmembrane  248 - 264 ( 242 - 267)
   INTEGRAL    Likelihood = -5.15   Transmembrane   75 -  91 (  74 -  94)
   INTEGRAL    Likelihood = -3.19   Transmembrane  110 - 126 ( 110 - 127)
   INTEGRAL    Likelihood = -2.13   Transmembrane  141 - 157 ( 138 - 157)
```

```
   INTEGRAL     Likelihood = -0.32   Transmembrane  188 - 204 ( 188 - 204)

----- Final Results -----
                bacterial membrane --- Certainty=0.5118(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06643 GB:AP001517 maltose/maltodextrin transport system
            (permease) [Bacillus halodurans]
 Identities = 117/281 (41%), Positives = 169/281 (59%), Gaps = 5/281 (1%)

Query: 1   MNKK--KRLNLTFVYILLIVLSIMWLFPIVWVVLTSFRGESSAFVNYFIPKTWTLDNYAK 58
            MNKK   RL +T +Y+ L+V+ I+ L+P++W V  S     S F +  IP+T +  +Y
Sbjct: 1   MNKKVKSRLEVTAIYLFLLVWGIVILYPLLWTVGLSLNEGTSLFSSRMIPETISPRHYEW 60

Query: 59  LFTQNTPFPFQQWFLNTLFVATCTCILSTLITVAMAYSLSRIKFKHRNGFLKLALVLNMPP 118
            LF      + QW+ NTL VA+ T + ST +     AY+ SR +F  R   L   L+L MPP
Sbjct: 61  LFFDPRSNYLQWYQNTLIVASVTSVCSTFLVALTAYAFSRYRFVGRTYGLYGLLLQMPP 120

Query: 119 GFMSMIAVYYILKALNLDQTLTALIFVY-SAGAALTFYIAKGFFDTIPYSLDESAMIDGA 177
            M+M+A+Y +L  +NL  TL  LI +Y     +  ++ KG+FDTIP  LDESA +DGA
Sbjct: 121 VLMAMVALYILLNTVNHLLDTLLGLILIYVGTSIPMNAFLVKGYFDTIPRELDESAKLDGA 180

Query: 178 TRLDIPLKITLPLSKPIIVVTALIAFMGPWNDFIFAKVILGDATSKYTVAIGLFSMLQQD 237
            IP  I LPL+KPI+   AL  FM P+MDFI  ++IL  + YT+A+GLF+ +
Sbjct: 181 GHSRIFFTIMLPLAKPILAVVALFNFMSPFMDFILPRILI-RSPENYTLALGLFNFVNDQ 239

Query: 238 TINQWFMSFTAGSVIIAIPITILFMFMQKYYVEGITGGSVK 278
              N  F  F AG+++IAIPI  +F+F+Q+Y + G+T G+ K
Sbjct: 240 FANN-FTRFAAGAILIAIPIATVFLFLQRYLISGLTTGATK 279
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5069> which encodes the amino acid sequence <SEQ ID 5070>. Analysis of this protein sequence reveals the following:

```
      Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL     Likelihood = -6.42   Transmembrane   76 -  92 (  71 -  97)
   INTEGRAL     Likelihood = -6.05   Transmembrane  248 - 264 ( 242 - 267)
   INTEGRAL     Likelihood = -3.50   Transmembrane  110 - 126 ( 110 - 127)
   INTEGRAL     Likelihood = -1.33   Transmembrane  129 - 145 ( 129 - 145)
   INTEGRAL     Likelihood = -1.33   Transmembrane  188 - 204 ( 188 - 204)

----- Final Results -----
                bacterial membrane --- Certainty=0.3569(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA60005 GB:X86014 cymG [Klebsiella oxytoca]
 Identities = 119/270 (44%), Positives = 172/270 (63%), Gaps = 7/270 (2%)

Query: 11  LVYATLIILSIIWLFPIAWVILTSFRSEGTAYVNYFIPKTFTLNHYINLFTNSTFPFGKW 70
            LVY  L++ +++ L P+ W +++S +     + + F   +FTL HY NL T    P+ KW
Sbjct: 12  LVYLFLLLNALVVLSPVTWTVMSSLKPGNNLFSGGFTEISFTLEHYHNLLTGT--PYLKW 69

Query: 71  FWNTLIVATFTCIISTFITVAIAYSLSRIKFRNGFLKLALLNMFPGFMSMIAIYYIL 130
            + NT I+AT   +IS  +   A+  SR +FK +   L  L+L MPP F+SM AIY +L
Sbjct: 70  YKNTFILATCNMLISLVVVTITAFISRYRFKAKKKILMSILVLQMFPAFLSMTAIYILL 129

Query: 131 KALGLTQTLTALVLVYSSGAALGF--YIAKGFFDTIPYSLDESAMIDGATRMDIFFKITL 188
             + L  T   L+LVY +G+ L F  ++ KG+FD IP SLDE+A IDGA  + IFF+I L
Sbjct: 130 SKMNLIDTYIGLLLVYVTGS-LPFMTWLVKGYFDAIPTSLDEAAKIDGAGHLTIFFEIIL 188

Query: 189 PLAKPIIVYTALLAFMGPWIDFIFAQVILGDATSKYTVAIGLFSMLQPDTINNWFMAFTA 248
```

```
            PLAKPI+V+ AL++F GPW+DFI   +IL  +  K T+AIG+FS +  ++  N F  F A
Sbjct: 189 PLAKPILVFVALVSFTGPWMDFILPILIL-RSEDKMTLAIGIFSWISSNSAEN-FTLFAA 246

Query: 249 GSVLIAVPITLLFMFMQKYYVEGITGGSVK 278
           G++L+AVPITLLF+  QK+   G+  G+VK
Sbjct: 247 GALLVAVPITLLFIVTQKHITTGLVSGAVK 276
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 227/278 (81%), Positives = 253/278 (90%)

Query: 1   MNKKKRLNNLTFVYILLIVLSIMMLFPIVWVVLTSFRGESSAFVNYFIPKTWTLDNYAKLF 60
           M  K+R  L  VY  LI+LSI+WLFPI WV+LTSFR E++A+VNYFIPKT+TL++Y  LF
Sbjct: 1   MKNKRRFQLSLVYATLIILSIIWLFPIAWVILTSFRSEGTAYVNYFIPKTFTLNHYINLF 60

Query: 61  TQNTFPFGQWFLNTLFVATCTCILSTLITVAMAYSLSRIKFKHRNGFLKLALVLNMFPGF 120
           T  TFPFG+WF+NTL VAT TCI+ST  ITVA+AYSLSRIKFK RNGFLKLALVLNMFPGF
Sbjct: 61  TNETFPFGKWFMNTLIVATFTCIISTPITVAIAYSLSRIKFKRRNGFLKLALILNMFPGF 120

Query: 121 MSMIAVYYILKALNLDQTLTALIFVYSAGAALTFYIAKGFFDTIPYSLDESAMIDGATRL 180
           MSMIA+YYILKAL L QTLTAL+ VYS+GAAL FYIAKGFFDTIPYSLDESAMIDGATR+
Sbjct: 121 MSMIAIYYILKALGLTQTLTALVLVYSSGAALGFYIAKGFFDTIPYSLDESAMIDGATRM 180

Query: 181 DIFLKITLPLSKPIIVYTALIAFMGPWMDFIFAKVILGDATSKYTVAIGLFSMLQDTIN 240
           DIF KITLPL+KPIIVYTAL+AFMGPW+DFIFA+VILGDATSKYTVAIGLFSMLQ DTIN
Sbjct: 181 DIFSKITLPLAKPIIVYTALLAFMGPWIDFIFAQVILGDATSKYTVAIGLFSMLQPDTIN 240

Query: 241 QWFMSFTAGSVIIAIPITLLFMFMQKYYVEGITGGSVK 278
            WFM+FTAGSV+IA+PIT+LFMFMQKYYVEGITGGSVK
Sbjct: 241 NWFMAFTAGSVLIAVPITLLFMFMQKYYVEGITGGSVK 278
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1640

A DNA sequence (GBSx1735) was identified in *S.agalactiae* <SEQ ID 5071> which encodes the amino acid sequence <SEQ ID 5072>. This protein is predicted to be cymF protein (malF). Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.46   Transmembrane  427 - 443 ( 417 - 447)
   INTEGRAL    Likelihood =-10.24   Transmembrane   99 - 115 (  96 - 121)
   INTEGRAL    Likelihood = -9.39   Transmembrane  166 - 182 ( 154 - 185)
   INTEGRAL    Likelihood = -6.21   Transmembrane  259 - 275 ( 257 - 276)
   INTEGRAL    Likelihood = -6.21   Transmembrane  229 - 245 ( 223 - 247)
   INTEGRAL    Likelihood = -6.10   Transmembrane   44 -  60 (  40 -  66)
   INTEGRAL    Likelihood = -4.51   Transmembrane  314 - 330 ( 312 - 331)

----- Final Results -----
            bacterial membrane --- Certainty=0.5585(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA60005 GB:X86014 cymF [Klebsiella oxytoca]
 Identities = 174/428 (40%), Positives = 263/428 (60%), Gaps = 21/428 (4%)

Query: 27  SFLIMGLANLKNKQIVKGLLFLISILPLFLTFVYQVIPAVKGLISLGTQEQGMTTKTVDG 86
           SFLIMG   L +   +KG +FL+ +I+ +I+ +  ++ A +GLI+LGT  Q     T  G
Sbjct: 15  SFLIMGATQLISGHWIKGSVFLLPQIV-VISNINLLLNATQGLITLGTVAQ-----TRSG 68

Query: 87  IKIQVATQGDNSMLMLIPGLASLIFCCVFAYIYWSNIKSAAHLLTLKEEGREIPSFKKDI 146
```

```
                I     GDNS+ ML+ G+ + IF     ++YW NIK A       +      SF + +
Sbjct: 69  FDI---VAGDNSIPMLVEGVVAFIFLFFSIFVYWLNIKDAQVCEKCHQ------SFTEQL 119

Query: 147 KSLTDGRFHMTLMSIPLIGVLLFTILPLVYMICLAPTNYDH-NHLPPKSLFDWVGFANFG 205
           +++ D RF   +++    I  + F I+P++  + ++ TNY   +H+PPK+L DWVG  NF
Sbjct: 120 RTIYDNRFATIMLAPAFIACIAFIIMPMIITVLVSLTNYSAPHHIPPKNLVDWVGLKNFI 179

Query: 206 NIPSGRMAS-TFFPVLSWTLIWAVFATVTNFFPGIIALLLINTKGLKPKMWRTIFVITM 264
            +F  R+ S TF  +  WT++WA FAT+    PG +LAL +  K +  KK WR +F++
Sbjct: 180 TLFELRIWSKTFVGIGVWTVLWAFFATLCTCSFGFLLALAIENKKIIAKKAWRVVFILFY 239

Query: 265 AVPQFISLLIMRNLLSDAGPVNALLIKWGLISSAHPLPFLSDPVWAKPSIIFVNMVGIP 324
            A+P F++LLI R LL+  GPVN+ L  WG+ S    + FLSDP+ AK ++I V++WVG P
Sbjct: 240 AIPAFVTLLIFRLLLNGIGPVNSTLNSWGIDS----IGFLSDPLIAKMTVIAVSVWVGAP 295

Query: 325 VTMLVATGIIMNLPAEQISAAEIDGANKFQVFQSITFPQILLIMTPTLIQQFIGNINNFN 384
             ML+ TG + N+P +  EA+E+DGA+KFQ P+ IT P +L  + P+L+  F  N NNF
Sbjct: 296 YPMLLITGANTNIPRDLYEASEVDGASKFQQFRETTLPMVLHQVAPSLVMTFAHNFNNFG 355

Query: 385 VIYLLTQGGPTNSTYYQAGSTDLLVTWLYNLTVTAADYNLASVVGILLFILSAVFSLLAY 444
            IYLLT+GGP N  Y  AG TD+L+TW+Y LT+    Y +ASV+ I+IP+  ++P++  +
Sbjct: 356 AIYLLTEGGPINPEYRFAGHTDILTTWIYKLTLDPQQYQIASVISIIIPLFLSIFAIWQF 415

Query: 445 TRTNSYKE 452
             R  S+KE
Sbjct: 416 RRMKSFKE 423
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5073> which encodes the amino acid sequence <SEQ ID 5074>. Analysis of this protein sequence reveals the following:

```
      Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.93    Transmembrane    98 - 114 (  95 - 122)
   INTEGRAL    Likelihood = -9.55    Transmembrane   165 - 181 ( 152 - 184)
   INTEGRAL    Likelihood = -9.24    Transmembrane   424 - 440 ( 419 - 443)
   INTEGRAL    Likelihood = -7.91    Transmembrane    43 -  59 (  39 -  71)
   INTEGRAL    Likelihood = -7.59    Transmembrane   258 - 274 ( 256 - 275)
   INTEGRAL    Likelihood = -6.21    Transmembrane   228 - 244 ( 222 - 246)
   INTEGRAL    Likelihood = -4.09    Transmembrane   311 - 327 ( 309 - 328)

----- Final Results -----
               bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA60005 GB:X86014 cymP [Klebsiella oxytoca]
 Identities = 179/426 (42%), Positives = 266/426 (62%), Gaps = 19/426 (4%)

Query: 26  SSIIMGFANFANKQFIKGILFLISELIPLVAPVSQIIPAIRGLVTLGYQTQGMTTKTIDG 85
           S +IMG     + +IKG +FL+ +++ +++ ++ ++ A +GL+TLGT  Q     T  G
Sbjct: 15  SFLIMGATQLISGHWIKGSVFLLFQIV-VISINILLNATQGLLTLGTVAQ-----TRSG 68

Query: 86  INIQVAVDCDNSMLMLIFGLASLIFCLVFAYIYWCNLKSARNLYLFKQKGQKIPSKEDL 145
            +I   V GDNS+ ML+ G+ + IF     ++YW N+K A+      Q     SF K L
Sbjct: 69  FDI---VAGDNSIPMLVEGVVAFIFLFFSIFVYWLNIKDAQVCEKCHQ------SFTEQL 119

Query: 146 ATLTNGRFHMTLMAIPLIGVLLFTILPLVYMICLAPTNFDH-NHLPPKSLFDWVGLANFG 204
             T+ + RF   ++A   I  + F I+P+I  + ++ TN+    +H+PPK+L DWVGL NF
Sbjct: 120 RTIYDNRFATIMLAPAFIACIAFIIMPMIITVLVSLTNYSAPHHIPPKNLVDWVGLKNFI 179

Query: 205 NVLSGRM-AGTFFPIFSWTLIWAVFATVTNFFPGIILALLINTKGLKWKKMWRTIFVITI 263
            +   R+ + TF  I  WT++WA FAT+    PG +LAL +  K +  KK WR +F++
Sbjct: 180 TLFELRIWSKTFVGIGVWTVLWAFFATLCTCSFGFLLALAIENKKIIAKKAWRVVFILFY 239

Query: 264 AVPQFISLLIMRNLLNDEGPLNALLNKIGLINGSLPFLSDPLWAKPSIIFVNMVGIPPFT 323

Query: 264 AVPQFISLLIMRNLLNDEGPLNALLNKIGLINGSLPFLSDPLWAKPSIIFVNMVGIPPFT 323
```

```
            A+P  F++LLI  R  LLN   GP+N+  LN   G+    S+  FLSDPL  AK  ++I  V++W+G  P+
Sbjct: 240  AIPAFVTLLIFRLLLNGIGPVNSTLNSWGI--DSIGFLSDPLIAKNTVIAVSVWVGAPYF  297

Query: 324  MLIATGIIMNLPSEQIRAARIDGASKFQVFKSITFPQILLIMTPNLIQQFIGNINNFNVI  383
            ML+  TG  +  N+P  +  EA+E+DGASKFQ  F+  IT  P  +L   +  P+L+   F   N  NNF   I
Sbjct: 298  MLLITGAMTNIPRDLYEASEVDGASKFQQFRHITLPMVLHQVAPSLVMTFAHNFNNFGAI  357

Query: 384  YLLITGGPTNSEVYQAGTTDLLVTNLYKLTVTAADYNLASVIGILIFVVSAIPSLLAYTR  443
             YLLT GGP N EY  AG TD+L+TW+YKLT++     Y +ASVI I+IF   +IF++  + R
Sbjct: 358  YLLTEGGPINPEYRFAGFTDLILITWIYKLTDFQQYQIASVISIIIFLFSIIFAIWQFRR  417

Query: 444  TASYKE  449
               S+KE
Sbjct: 418  MKSFKE  423
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 357/446 (80%), Positives = 404/446 (90%), Gaps = 2/446 (0%)

Query: 11   MSLKEVFQKGDLATKLSFLIMGLANLKNKQIVKGLLFLISELLFLITFVYQVIPAVKGLI  70
            +S+ E  ++G    KLS +IMG AN  NKQ +KG+LFLISE++FL+ FV Q+IPA++GL+
Sbjct: 10   ISVIEALKRGSWDIKLSSIIMGFANFANKQFIKGILFLISELIFLVAFVSQIIPAIRGLV  69

Query: 71   SLGTQRQGMTTKTVDGIKIQVATQGDNSMLMLIFGLASLIFCCVPAYIYWSNIKSAHLL  130
            +LGTQ QGMTTKT+DGI IQVA  GDNSMLMLIFGLASLIFC VPAYIYW N+KSA +L
Sbjct: 70   TLGIQTQGMTTKTIDGINIQVAVDGDNSMLMLIFGLASLIFCLVPAYIYWCNLKSARNLY  129

Query: 131  TLKEEGRRIPSFKKDIKSLTDGRFHMTLMSIPLIGVLLFTILPLVYMICLAFTNYDENHL  190
              K++G++IPSFK+D+ +LT+GRFHMTLM+IPLIGVLLFTILPL+YMICLAFTN+DENHL
Sbjct: 130  LFKQKGQKIPSFKEDLATLTNGRFHMTLMAIPLIGVLLFTILPLIYMICLAFTNFDENHL  189

Query: 191  PPKSLFDWVGFANFGNIFSGRMASTFFPVLSWTLIWAVPATVINFFFGIILALLINTKGL  250
             PPKSLFDWVG ANFGN+ SGRMA TFFP+ SWTLIWAVPATVINFFFGIILALLINTKGL
Sbjct: 190  PPKSLFDWVGLANFGNVLSGRMAGTFFPIFSWTLIWAVPATVINFFFGIILALLINTKGL  249

Query: 251  KFKKMWRTIFVITMAVPQFISLLIMRNLLSDAGPVNALLKWGLISSAHPLPFLSDPVMA  310
            K+KMWRTIFVIT+AVPQFISLLIMRNLL+D GP+NALL K GLI+ +  LPFLSDP+MA
Sbjct: 250  KWKKMWRTIFVITIAVPQFISLLIMRNLLNDEGPLNALLNKIGLINGS--LPFLSDPLMA  307

Query: 311  KPSIIFVNMWVGIPVTMLVATGIIMNLPAEQIEAAEIDGANKFQVFQSITFPQILLIMTP  370
            KPSIIFVNMW+GIP TML+ATGIIMNLP+EQIEAAEIDGA+KFQVF+SITFPQILLIMTP
Sbjct: 308  KPSIIFVNMWIGIPFTMLIATGIIMNLPSEQIEAAEIDGASKFQVFKSITFPQILLIMTP  367

Query: 371  TLIQQFIGNINNFNVIYLLTGGPTNSTYYQAGSTDLLVTWLYNLTVTAADYNLASVVGI  430
              LIQQFIGNINNFNVIYLLT GGPTNS YYQAG+TDLLVTWLY LTVTAADYNLASV+GI
Sbjct: 368  NLIQQFIGNINNFNVIYLLTGGGPTNSEYYQAGTTDLLVTWLYKLTVTAADYNLASVIGI  427

Query: 431  LIFILSAVFSLLAYTRTNSYKEGAAK  456
             LIF +SA+FSLLAYTRT SYKEGAAK
Sbjct: 428  LIFTVSAIFSLLAYTRTASYKEGAAK  453
```

A related GBS gene <SEQ ID 8869> and protein <SEQ ID 8870> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:    -12.73
GvH: Signal Score (-7.5): -6.04
     Possible site: 36
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 7 value: -11.46 threshold:  0.0
   INTEGRAL    Likelihood =-11.46   Transmembrane  427 - 443 ( 417 - 447)
   INTEGRAL    Likelihood = -9.87   Transmembrane   99 - 115 (  96 - 121)
   INTEGRAL    Likelihood = -9.39   Transmembrane  166 - 182 ( 154 - 185)
   INTEGRAL    Likelihood = -6.21   Transmembrane  259 - 275 ( 257 - 276)
   INTEGRAL    Likelihood = -6.21   Transmembrane  229 - 245 ( 223 - 247)
   INTEGRAL    Likelihood = -6.10   Transmembrane   44 -  60 (  40 -  66)
   INTEGRAL    Likelihood = -4.51   Transmembrane  314 - 330 ( 312 - 331)
```

```
  PERIPHERAL  Likelihood =  0.90     212
 modified ALOM score:   2.79

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.5585(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01027(379 - 1656 of 1968)
EGAD|33392|34706(15    -    423    of    427)    cymF    protein    {Klebsiella    oxytoca}
GP|854233|emb|CAA60005.1|X86014 cymF {Klebsiella oxytoca} PIR|S63615|S63615 malF protein
homolog cymF - Klebsiella oxytoca
%Match = 23.8
%Identity = 41.3  %Similarity = 64.5
Matches = 171  Mismatches = 140  Conservative Sub.s = 96

132       162       192       222       252       282       312       342
VLLFLAILTVVKSNLAITLNV*NNSIKTSLKQNSTSRVMR*GEYSSFQLRVLPISYFVK*QLKETIMNKKLISLDGMSLK

                                                                                ML

372       402       432       462       492       522       552       582
EVPQKGDLATKLSFLIMGLANLKNKQIVKGLLFLISEILFLITFVYQVIPAVKGLISLGTQEQGWITKTVDGIKIQVATQ
                 ||||||    | : : :|| :||: :|  :|: :  :: | :|||:||| |     |  |  |
LSEKGKSMRIFPASFLIMGATQLISGHWIKGSVFDLFQI-VVISNINLLLNATQGLITVLGTVAQ-----TRSGFDI---VA
                 20        30        40        50        60                70

612       642       672       702       732       762       792       822
GDNSMLMLIFGLASLIFXCVFAYIYWSNIXSAAHLLTLKEEGREIPSFKKDIKSLTGRPHMTLMSIPLIGVLLFTILPL
||||::||: |: ::||     ::|| ||          :  :    || : :::: | ||   :::  :|  : | |:|:
GDNSIPKMLVEGVVAFIFLFFSIFVYWLNI------KDAQVCEKCHQQSFTEQLRTIYDNRFATIMLAPAPLACIAFIIMPM
                 90        100               110       120       130       140

852       879       909       939       966       996       1026      1056
VYMICLAPTNYDH-NHLPPKSLPDWVGFANPGNIPSGRMAS-TFFPVLGWTLIWAVPATVTNFFPGIILALLINTKGLKF
:  : :::||   :|:|||:| ||||: ||  :|  |: | ||  :  ||::|| |||:    || :||| :  | :
IITVLVSLTNYSAPHHIPPKNLVDWVGLKNFITLPELRIWSKTFVGIGVWTVLWAFFATLCTCSFGFLLALALENKKIIA
                 160       170       180       190       200       210       220

1086      1116      1146      1176      1206      1236      1266      1296
KGWRTIPVITWAVPQFISLLIMRNLLSDAGFVNALLIKWGLISSAHLPFLSDFVWAKFSIIFVNMWGIPVTMLVATG
|| || :|::  |:| |::||| | ||:  ||||: |  ||       : |||||: || ::| |::||| |  ||: ||
KVARNVVFILPYAIPAFVTLLIFRLLLNGIGFVNSTLNSWG----IDSIGFLSDFLIAKMTVIAVSVWVGAPYFALLITG
                 240       250       260               270       280       290       300

1326      1356      1386      1416      1446      1476      1506      1536
IINNLPABQIEAAEIDCANKFQVPQSITFFQILLIMTPTLIQQFIGNINNFNVIYLLTQGGPTNSTYYQAGSTDLLVTWL
 : |: :  |:|:|||:|'| |: ||:| :|  : |:|:  |  | |||  ||||:||| |  |  || ||:|:||:
AMTNIPRDLYEASEVDCASKFQQFREITLPMVLHQVAPSLVMTFAHNFNNFGAIYLLTEGGPINPEYRFAGHTDILITWI
                 320       330       340       350       360       370       380

1566      1596      1626      1656      1686      1716      1746      1776
YNLTVTAADYNLASVVGILIFILSAVPSLLAYTRTNSYKEGAAK**IRKNVLTLLLFIFY**YYQLCGSFPLFGSFSQAS
| ||:    | :|||: |:||:: ::|:: : |  |:||
YKLTLDPQQYQIASVISIIIFLFLSIFAIWQPREMKSFKEDVGM
                 400       410       420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1641

A DNA sequence (GBSx1736) was identified in *S.agalactiae* <SEQ ID 5075> which encodes the amino acid sequence <SEQ ID 5076>. This protein is predicted to be maltose/maltodextrin-binding protein precursor. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.98   Transmembrane   25 -  41 (  24 -  43)

----- Final Results -----
                bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9999> which encodes amino acid sequence <SEQ ID 10000> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA26925 GB:L08611 MalX [Streptococcus pneumoniae]
 Identities = 117/418 (27%), Positives = 186/418 (43%), Gaps = 43/418 (10%)

Query: 15  TKMSKVTWKKLLVSTAALSVVAGAIAATHSNSVDAASKTTIKLWVPTDSKASYKAIVKK 74
           +K  K+T     V+ A+L +VA G+  A        ++    + ++V    K+  + + K
Sbjct: 3   SKFMKSTAVLGTVTLASLLLVACGSKTADKPADSGSSEVKELTVTVDEGYKSYIEEVAKA 62

Query: 75  FZKE-NKGVTVKMIESNDSKAQENVKKDPSKAADVFSLPHDQLGQLVESGVIQRIPEQYS 133
           ++KE    VT+K  ++     + ++         DV   P+D++G L   G + E+  + S
Sbjct: 63  YEKEAGVKVTLKTGDALGGLDKLSLDNQSGNVPDVVMAPYDRVGSLGSDGQLSEV--KLS 120

Query: 134 KSIAKNDTKQSLTGAQYKGKTYAFPFGIESQVLYYNKTKLTADDVKSYETITSKGKFGXQ 193
               +DT +SL  A   GK Y  P  IES V+YYNK  L  D  K++  + +  K
Sbjct: 121 DGAKTDDTKSLVTAA-NGKVYGAPAVIESLVMYYNKD-LVKDAPKTFADLENLAKDSKY 178

Query: 194 LKAA-----------NSYVTGPXFLVGDTLFGKSGEDAKGTWGNEAGVSVL----- 235
            A               N Y T     G  +FG++G+DAK     N+  ++ +
Sbjct: 179 AFAGEDGKTTAFLADWTNFYYTYGLLAGNGAYVFQQNGKDAKDIGLANDGSIAGINYAKS 238

Query: 236 ---KWIADQKNDGFVNLTAENTMSKFGDGSVHAFESGPWDYDAAKKAVGEDKIGVAVYP 292
               KW    +G  NL      ++P +G   A   GPW   A K A  +   GVA  P
Sbjct: 239 WYEKWPKGMQDTEGAGNLI----QTQPQGKTAAIIDGPWKAQAFKDA--KVNYGVATIP 292

Query: 293 TMKIGDKEVQQKAFLGVKLYAVNQAPAGSNTKRISASYKLAAYLTNAESQKIQFEKRHIV 352
           T+  G    +  AF G K + + QA       K + AS K   +L   E QK+ ++K + +
Sbjct: 293 TLPNGK---EYAAFGGGKAWVIPQA-----VKNLEASQKFVDFLVATEQQKVLYDKTNEI 344

Query: 353 PANSSIQSSDSVQKDELAKAVIEWGSSDKYTTVMPKLSQMSTFWTESAAILSDTYSGK 410
           PAN+  +S    + DEL  AVI+      K T  +P +SQMS  W  +  +L D  SG+
Sbjct: 345 PANTEARSYAEGKNDELTTAVIK---QFKNTQPLPNISQMSAVWDPAXNMLFDAVSGQ 399
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5077> which encodes the amino acid sequence <SEQ ID 5078>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> May be a lipoprotein

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAA26925 GB:L08611 MalX [Streptococcus pneumoniae]
 Identities = 126/423 (29%), Positives = 191/423 (44%), Gaps = 50/423 (11%)

Query: 13  SLTLASTLLVAGCSGSGSKDK--KKAGADSKTIKLWVPTGSKKSYADTIAK-PKKDSGYTVK 69
           ++TLAS LLV CGS + DK       ++ K + ++V  G  KSY + +AK +EK++G  V
Sbjct: 14  TVTLASLLLVACGSKTADKPADSGSSKVKELTYVYDKG-YKSYIEEVAKAYEKEAGVKVT 72

Query: 70  VVESEDPKAQEKIKKD--ASTAADVFSLPHDQLGQLVESGTIQEVPEKYNKEIAATSTDQ 127
           +   +    +K+  D  +     DV   P+D++G L   G + EV  K +       T +
Sbjct: 73  LKTGDALGGLDKLSLDNQSGNVPDVMAPYDRVGSLGSDGQLSEV--KLSDGAKTDDTTK 130

Query: 128 ALVGAQYKGKTYAFPFGIESQVLFYNKSKLAAEDVTSYD----TTTTKATGGTFKQ--- 180
           +LV A   GK Y  P  IES V++YNK  +     T  D          +K  F G   +
Sbjct: 131 SLVTAA-NGKVYGAPAVIESLVMYYNKDLVKDAPKTFADLENLAKDSKYAFAGEDGKTTA 189

Query: 181 -----ANTYATGPLFMSVGNTLFGENGEDVKGTNWGNEKGAAVL--------KWIADQAS 227
                N Y T  L   G  +FG+NG+D K     N+   A +         KW
Sbjct: 190 FLADWTNFYTTYGLLAGNGAYVFGQNGKDAKDIGLANDGSIAGINYAKSWYEKWPKGMQD 249

Query: 228 NKGFVSLDANNVMSKFGDGSVASFESGPWDYEAAQKAIGKENLGVAIYPKVTIGGETVQQ 287
           +G      N + ++P +G  A+   GPW  +A + A  K N GVA   P +  G E
Sbjct: 250 TEG----AGNLIQTQFQQGKIAAIIDGPWAQAFKDA--KVNYGVATIPTLPNGKE---Y 300

Query: 288 KAFLGVKLYAVNQAPAKGDTKRIAASYKLASYLTNAESQENQFKTRNIVPANKEVQSSEA 347
            AF G K + + QA      K + AS K   +L   E Q+  +   N +PAN E +S
Sbjct: 301 AAFGGGKAWVIPQA-----VKNLEASQKFVDFLVATEQQKVLYDKTNEIPANTEARSYAE 355

Query: 348 VQSNELAKTVITMGSSSDYTVVMPKLSQMGTFWTESAAILSDAPNG----KIKENDYLTK 403
           +++EL  VI   +  T  +P +SQM  W  + +L DA +G    K   ND +T
Sbjct: 356 GKNDELTTAVIKQPKN---TQPLFNISQMSAVWDPAKNMLFDAVSGQKDAKTAANDAVTL 412

Query: 404 LQQ 406
           +++
Sbjct: 413 IKE 415
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 278/415 (66%), Positives = 334/415 (79%), Gaps = 6/415 (1%)

Query: 21  TWKKLLVSTAALSVVAGGAIAATHSNSVD----AASKTTIKLWVPTDSKASYKAIVKKFZ 76
           +W+K++V  A+L++ A   +   S S D     A  TIKLWVPT SK SY   + KF+
Sbjct: 3   SWQKVIVGGASLTL-ASTLLVGCSGSGSKDKKEAGADSKTIKLWVPTGSKKSYADTIAKFE 61

Query: 77  KENKGVTVKMIESNDSKAQENVKKDPSKAADVFSLPHDQLGQLVESGVIQEIPEQYSKEI 136
           K++ G TVK++ES D KAQE +KKD S AADVFSLPHDQLGQLVESG IQE+PE+Y+KEI
Sbjct: 62  KDS-GYTVKVVESEDPKAQEKIKKDASTAADVFSLPHDQLGQLVESGTIQEVPEKYNKEI 120

Query: 137 AKNDTKQSLTGAQYKGKTYAFPFGIESQVLFYNKTKLTADDVKSYETTTSEGKFGKQLKA 196
           A  T Q+L GAQYKGKTYAFPFGIESQVLFYNK+KL A+DV SY+TIT+K  FG  K
Sbjct: 121 AATSTDQALVGAQYKGKTYAFPFGIESQVLFYNKSKLAAEDVTSYDTITTKATFGGTFKQ 180

Query: 197 ANSYVTGPXPLGVGDTLFGKSGEDAKGTNWGNEAGVSVLKWIADQKNDGFVNLTAENTM 256
           AN+Y TGP F+SVG+TLFG+ +GED KGTNWGNE G +VLKWIADQ  N GFV+L A N M
Sbjct: 181 ANTYATGPLFMSVGNTLFGENGEDVKGTNWGNSKGAAVLKWIADQASNKGFVSLDANNVM 240

Query: 257 SKFGDGSVHAFESGPWDYDAAKKAVGEDKIGVAVYPTMKIGDKEVQQKAFLGVKLYAVNQ 316
           SKFGDGSV +FESGPWDY+AA+KA+G++ +GVA+YP + IG + VQQKAFLGVKLYAVNQ
Sbjct: 241 SKFGDGSVASFESGPWDYEAAQKAIGKENLGVAIYPKVTIGGETVQQKAFLGVKLYAVNQ 300

Query: 317 APAGSNTKRISASYKLAAYLTNAESQKIQPEKRHIVPANSSIQSSDSVQKDELAKAVIEM 376
           APA  +TKRI+ASYKLA+YLTNAESQ+ QP+ R+IVPAN  +QSS++VQ +ELAK VI M
Sbjct: 301 APAKGDTKRIAASYKLASYLTNAESQENQFKTRNIVPANKEVQSSEAVQSNELAKTVITM 360

Query: 377 GSSDKYTTVMPKLSQMSTFWTESAAILSDTYSGKIKSSDYLKRKLQFDKDIAKTK 431
           GSS  YT VMPKLSQM TFWTESAAILSD ++GKIK +DYL +L+QFDKDIA TK
Sbjct: 361 GSSDYTVVMPKLSQMGTFWTESAAILSDAPNGKIKENDYLTKLQQFDKDIAATK 415
```

SEQ ID 5076 (GBS649) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 2 & 3; MW 76kDa) and in Figure 186 (lane 7; MW 76kDa).. It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 7; MW 51kDa) and in Figure 178 (lane 8; MW 51kDa).

GBS649-His was purified as shown in Figure 229, lane 8. Purified GBS649-GST is shown in Figure 245, lanes 6 &73.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1642

A DNA sequence (GBSx1737) was identified in *S.agalactiae* <SEQ ID 5079> which encodes the amino acid sequence <SEQ ID 5080>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.2462(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD02112 GB:AF039082 putative maltose operon transcriptional
          repressor [Lactococcus lactis]
 Identities = 43/61 (70%), Positives = 49/61 (79%)

Query: 2  VTIKDVAAKAGVNPSTVSRVLKDNASISSKTKERVKKAMEELGYVPNVAAQMLASGLTQN 61
          VTIKDVA KAGVN STVSRV+KD++ IS KTK +V+KAM ELGY  N AAQ+LASG T
Sbjct: 3  VTIKDVAKKAGVNASTVSRVIKDSSEISDKTKVKVRKAMEELGYRRNAAAQILASGKTNT 62

Query: 62 I 62
          I
Sbjct: 63 I 63
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5081> which encodes the amino acid sequence <SEQ ID 5082>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.93   Transmembrane  269 - 285 ( 266 - 287)

----- Final Results -----
                bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 53/62 (85%), Positives = 57/62 (91%)

Query: 1  MVTIKDVAAKAGVNPSTVSRVLKDNASISSKTKERVKKAMEELGYVPNVAAQMLASGLTQ 60
          MVTIKDVA KAGVNPSTVSRVLKDN SIS KTKE+V+KAM +LGYVPNVAAQ+LASGLT
Sbjct: 26 MVTIKDVACKAGVNPSTVSRVLKDNRSISMKTKEKVRKAMADLGYVPNVAAQILASGLTH 85

Query: 61 NI 62
          NI
Sbjct: 86 NI 87
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1643

A DNA sequence (GBSx1738) was identified in *S.agalactiae* <SEQ ID 5083> which encodes the amino acid sequence <SEQ ID 5084>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.70   Transmembrane   14 -  30 (   8 -  34)
   INTEGRAL    Likelihood = -6.90   Transmembrane   66 -  82 (  63 -  85)
   INTEGRAL    Likelihood = -6.69   Transmembrane  110 - 126 ( 105 - 128)
   INTEGRAL    Likelihood = -3.93   Transmembrane  132 - 148 ( 129 - 149)

----- Final Results -----
                bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9443> which encodes amino acid sequence <SEQ ID 9444> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC67260 GB:AF017113 YvjA [Bacillus subtilis]
 Identities = 83/227 (36%), Positives = 140/227 (61%)

Query: 9   PGWDSAPFIMINIPLLLLCYPGLGKQTFLKTVYGSWIFPVFIKLTQSVFTLTHNPLLAA 68
           +G+++A+    IINIPL +      LG +  LKT+ GS    P+ + LT+ +   TH+ LLAA
Sbjct: 52  YGFEAAYVQWIINIPLFIAGVILLGSKFGLKTLAGSVFLPLVVFLTRDIQPATHHELAAA 111

Query: 69  LFGGVIVGCGLGIVFWDSSSTGGTGIIIQFLGKYTPISLGQGVILIDGLVTIVGFLAFDS 128
           +FGG7 +G G+GIV+     STGGT +   Q + KY+ +SLG+ + +IDG++ +    + F+
Sbjct: 112 IFGGVGIGIGIGIVVLGKGSTGGTALAAQIIHKYSGLSLGKCLAIIDGMIVVTAMIVFNI 171

Query: 129 DTVMPSIIGLITISYIINAIQTGFTLSTVLIVSQEHQKIKTYINTVADRGVTEIPVKGG 188
           +  +++++G+   S  I+ +Q GF      LI+++ Q +K  +     DRGVT+I   GG
Sbjct: 172 EQGLYAMLSVVVSSKTIDVVQVGFNRSKMALIITKQSQAVKEAVLQKIDRGVTKISAVGG 231

Query: 189 YSGTNQIMLMTTAGYEPAKLQSAIAEIDETAFITVFTPTSQASGRGF 235
           Y+  ++ +IM  +   EF KL++ + +IDE+AF+ V   S+  G GF
Sbjct: 232 YTDDDRPILMCVVGQTEPTKLKQIVKQIDESAFVIVADASEVLGEGF 278
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5085> which encodes the amino acid sequence <SEQ ID 5086>. Analysis of this protein sequence reveals the following:

```
      Possible site: 57
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.21   Transmembrane  104 - 120 ( 101 - 123)
   INTEGRAL    Likelihood = -3.93   Transmembrane  147 - 163 ( 142 - 167)
   INTEGRAL    Likelihood = -3.29   Transmembrane  169 - 185 ( 169 - 186)

----- Final Results -----
                bacterial membrane --- Certainty=0.3484(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC67260 GB:AF017113 YvjA [Bacillus subtilis]
 Identities = 106/267 (39%), Positives = 169/267 (62%), Gaps = 1/267 (0%)

Query: 7   DLLLVTIGSFITAIGFNTMFVDNHIASGGMVGIAVVIKALPGISPSLPLMASNIPLLLMC 66
```

```
              D + + IG+ ITA+ FN   + N IA+GG+ GI+ ++++ +G   +       NIPL +
Sbjct: 13  DYVYILIGAAITAVSPNVFLLPNKIAAGGVSGISTILQS-VGFEAAYVQWIINIPLFIAG 71

Query: 67  YPFLGKQNPIKTLYGSWIYPIAIRSTNSLPTLTHNQLLAAIFGGIICGIGLGMVPWGNSS 126
              LG +  +KTL GS   P+ +   T +   TH++LLAAIFGG+  GIG+G+V+ G  S
Sbjct: 72  VILLGGKFGLKTLAGSVFLPLVVFLTRDIQPATHHRLLAAIFGGVGIGIGIGIVYLGKGS 131

Query: 127 TGGTGILTQILHKYSPLSLGVAMTIVDGISVLMGFIALSADDVMYSTIGLFVIGYVISVM 186
              TGGT +  QI+HKYS LSLG  + I+DG+ V+   I  + +  +Y+ +G++V     I V+
Sbjct: 132 TGGTALAAQIIHKYSGLSLGKCLAIIDGMIVVTAMIVFNIEQGLYAMLGVYVSSKTIDVV 191

Query: 187 ENGFDSSKNVMIISKDYQAIREYITTVMDRGVTKLPIRGGYTTSDKIMLMAIVSSHELPT 246
              + GF+ SK  +II+K  QA++E +   +DRGVTK+   GGYT  D+ +LM +V   E
Sbjct: 192 QVGPNRSKMALIITKQEQAVKEAVLQKIDRGVTKISAVGGYTDDDRPILMCVVGQTEFTK 251

Query: 247 LQEKILEIDDTAFIVVMPAAQVMGRGF 273
              L++ + +ID++AF++V  A++V+G GF
Sbjct: 252 LKQIVKQIDESAFVIVADASEVLSEGF 278
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 135/252 (53%), Positives = 190/252 (74%)

Query: 1   MAVSFHSVPGMDSAFFIMIINIPLLLLCYFGLGKQTFLKTVYGSWIFVVFIKLTQSVPTL 60
            +AV   +PG   + F+M  NIPLLL+CYF LGKQ F+KT+YGSWI+P+ I+ T S+PTL
Sbjct: 39  IAVVIKALFGISPSLFLMASNIPLLLMCYFFLGKQNFIKTLYGSWIYPIAIRSTNSLPTL 98

Query: 61  THNPLLAALFGGVIVGCGLGIVPWSDSSTGGTGIIIQFLGKYTPISLGQGVILIDGLVTI 120
            THN LLAA+FGG+I G GLG+VPW +SSTGGTGI+ Q L KY+P+SLG  + ++DG+  +
Sbjct: 99  THNQLLAAIFGGIICGIGLGMVPWGNSSTGGTGILTQILHKYSPLSLGVAMTIVDGISVL 158

Query: 121 VGFLAFDSDTVMFSIIGLITISYIINAIQTGFTTLSTVLIVSQEHQKIKTYINTVADRGV 180
            +GF+A  +D VM+S IGL  I Y+I+ ++ GF +   V+I+S+++Q I+ YI TV DRGV
Sbjct: 159 MGFIALSADDVMYSTIGLFVIGYVISVMENGFDSSKNVMIISKDYQAIREYITTVMDRGV 218

Query: 181 TEIPVKGGVSGTNQIMLMTTIAGYEFAKLQSAIAEIDETAPITVTPTSQASGRGFSLQKN 240
            T++P++GGV+ +++IMLM  ++ +E   LQE I EID+TAFI V P +O  GRGFSL K
Sbjct: 219 TKLPIRGGYTTSDKIMLMAIVSSHELPTLQEKILEIDDTAFIVVMPAAQVMGRGFSLTKQ 278

Query: 241 HGRLDEDILMPM 252
            + R D+D+L+PM
Sbjct: 279 YKREDKDVLLPM 290
```

A related GBS gene <SEQ ID 8871> and protein <SEQ ID 8872> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Cmend: 6
McG: Discrim Score:       1.57
GvH: Signal Score (-7.5): -2.56
     Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 4 value:  -7.70 threshold:  0.0
   INTEGRAL    Likelihood = -7.70   Transmembrane   14 -  30 (   8 -  34)
   INTEGRAL    Likelihood = -6.90   Transmembrane   66 -  82 (  63 -  85)
   INTEGRAL    Likelihood = -6.69   Transmembrane  110 - 126 ( 105 - 128)
   INTEGRAL    Likelihood = -3.93   Transmembrane  132 - 148 ( 129 - 149)
   PERIPHERAL  Likelihood =  3.71       37
 modified ALOM score:   2.04

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02139(118 - 1008 of 1356)
OMNI|NT01BS4111(51 - 325 of 327) conserved hypothetical protein
%Match = 19.3
%Identity = 37.1  %Similarity = 62.1
Matches = 101  Mismatches = 99  Conservative Sub.s = 68

27        57        87        117                 165
           ARAIPSFIVGSALTGALVGLAGIKLMAPHGGIFVIALTSNPLLYIL----FILIGAVVSGVLFGLF---
                                                |  ||:||  :|||| :: | | :|
VCFFISYILDFTAALAYYHCIWVLFTSNQQRIKMLSSSIGRNGGYMMDVRNKTLWILRDVVYILIGAAITAVSFNVFLLP
          10        20        30        40        50        60        70        80

216       246       276       306       336       366       396       426
RKIK*LISTYPNLR*IKGE*XIVILXXLIN*XXGGISGLAVSPKEVPGWDSAFPIMIINIPLLLLCYPGLGKQTFLKTVY
 ||                              ||:||::    : :|:::|:   ||||||::     || :  |||:
NKI-----------------------------AAGGVSGIST-ILQSYGFKAAYVQWIINIPLFTAGVILGGKFGLKTLA
                                  90        100       110       120       130

456       486       516       546       576       606       636       666
GSWIFPVFIKLTQSVPTLTHNPLLAALFGGVIVGCSLGIVFWSDSSTGGTNGIIIQFLGKYTPISLGQGVILIDGLVTIVG
||  :|: : ||: :  ||: ||||:|||| :| |:|||:     ||||  :  | :||: :|||: : :|||:: :
GSVFLPLVVFLTRDIQPATHHELLAAIFGGVGIGIGIGIVLGKGSTGGTALAAQIIKCYSGLSLGKCLAIIDGMLVTTA
                  150       160       170       180       190       200       210

696       726       756       786       816       846       876       906
FLAPDSDTVMPSIIGLITISYIINAIQTGFTTLSTVLIVSQEHQKIKTYINTVADRGVTEIPVKGGYSGTNQIMLMTTIA
 : |: :  :::::|:   |  |: :| ||      ||:::: | :|  :     |||||:|   |||:  :: :||  :
MIVFNIRQQLYAMLGVYVSSKTIDVVQVGFNRSKMALIITKQEQAVKSAVLQKIDRGVTKISAVGGYTDDDRPLIMCVVG
                  230       240       250       260       270       280       290

936       966       996       1026      1056      1086      1116
GYEFAKLQSAIAEIDETAPITVTPTSQASGRGFSLQKNHGRLDEDILMPM*SIDN*SFF**NSR*NIHKR*QNC
 || ||:: : :|||:||:|  |   |:  : ||
QTEFTKLKQIVKQIDESAPVIVADASEVLGSGFKRA
                  310       320
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1644

A DNA sequence (GBSx1739) was identified in *S.agalactiae* <SEQ ID 5087> which encodes the amino acid sequence <SEQ ID 5088>. This protein is predicted to be ABC transporter, ATP-binding protein (b0820). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3122(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC24918 GB:AF012285 YkpA [Bacillus subtilis]
 Identities = 355/540 (65%), Positives = 451/540 (82%), Gaps = 4/540 (0%)

Query: 1   MLTVSDVSLRFSDRKLFDEVNINFTAGNTYGLIGANGAGKSTFLKILAGDIRPTTGHIAL 60
            M+ V++VSLRF+DRKLF++VNI FT GN YGLIGANGAGKSTFLK+L+G+IRP TG + +
Sbjct: 1   MIAVNNVSLRFADRKLFEDVNIKFTPGNCYGLIGANGAGKSTFLKVLSGRIKPQTGDVHM 60

Query: 61  GPDERLSVLRQNHFDYSDERVIDVVIMGNETLYSIMKEKDAIYMKEDPSDEDGVRAAELE 120
            P EKL+VL+QNHF+YE+  V+ VVIMG++ LY +M+EKDAIYMK DPSDEDG+RAAELE
Sbjct: 61  SPGERLAVLKQNHFRYEEYEVLKVVIMGHKRKLYEVMQEKDAIYMKDPSDEDGIRAAELE 120
```

```
Query: 121 GEPAELGGWEAESEASQLLQNLNISEELHYQNMSELANGDKVKVLLAKALPGKPDVLLLD 180
           GEPAEL GWEAESEA+ LL+ L ISE+LH + M++L   +KVKVLLA+ALPGKPDVLLLD
Sbjct: 121 GEPAELNGWEAESEAAILLKGLISEDLHTKKMADLGGSEKVKVLLAQALPGKPDVLLLD 180

Query: 181 EPTNGLDIQSITWLEDFLIDFENTVIVVSHDRHFLNKVCTHMADLDFGKIKLFVGNYDFW 240
           EPTN LD+Q+I WLE+FLI+FENTVIVVSHDRHFLNEVCTH+ADLDF KI+++VGNYDFW
Sbjct: 181 EPTNHLDLQAIQWLKEFLINFENTVIVVSHDRHFLNKVCTHIADLDFNKIQIYVGNYDFW 240

Query: 241 KESSELAARLQADRNAKAEEKIKQLQEFVARFSANASKSKQATSRKKMLDKIELEEIVPS 300
            ESS+LA +L  + N K EE+IKQLQEFVARFSANASKSKQATSRKK+L+KI L++I PS
Sbjct: 241 YESSQLALKLSQEANKKEEQIKQLQEFVARFSANASKSKQATSRKKLLEKITLDDIKPS 300

Query: 301 SRKYPFVNFKAERKMGNDLLTVENLSVTIDGKKILDNISFILRPGDKTALIGQNDIQTTA 360
           SR+YP+VNF  ERE+GND+L VE L+ TIDG K+LDN+SFI+   DK A  G+N++  T
Sbjct: 301 SRRYPYVNFTPERKIGNDVLRVEGLTKTIDGVKVLDNVSFIMNREDKIAFTGRNELAVTT 360

Query: 361 LIRALMGDIEYE-GTIKWGVTTSRSYLPKDNSRDFASGE-SILEWLRQPASKEEDDNTFL 418
           L + + G++E + GT KWGVTTS++Y PKDNS  F   + +++WLRQ+ S  +   +FL
Sbjct: 361 LFKIISGEMEADSGTFKWGVTTSQAYFPKDNSEYFEGSDLNLVDWLRQY-SPHDQSESFL 419

Query: 419 RGFLGRMLPSGDEVNKSVNVLSGGEKVRVMLSKLMLLKSNVLVLLDEPTNHLDLESISSLN 478
           RGFLGRMLPSG+EV+K  NVLSGGEKVR MLSK ML  +N+L+LD+PTNHLDLESI++LN
Sbjct: 420 RGFLGRMLPSGEEVHKKANVLSGGEKVRCMLSKAMLSGANILILDEPTNHLDLESITALN 479

Query: 479 DGLKDFKESIIFASHDHEFIQTLANHIIVLSKNGVIDRIDETYDEFLENTEVQAKVAQLW 538
           +GL  FK +++F SHDH+F+QT+ANH II ++ NG++D+   +YDEFLEN +VQ K+ +L+
Sbjct: 480 NGLISFKGAMLFTSHDHQFVQTIANRIIEITPNGIVDK-QMSYDEFLENADVQKKLTELY 538
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5089> which encodes the amino acid sequence <SEQ ID 5090>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3124(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 497/539 (92%), Positives = 525/539 (97%)

Query: 1   MLTVSDVSLRFSDRKLFDEVNINFTAGNTYGLIGANGAGKSTFLKILAGDIEPTTGHIAL 60
           +LTVSDVSLRFSDRKLFD+VNI FTAGNTYGLIGANGAGKSTFLKILAGDIEP+TGHI+L
Sbjct: 1   LLTVSDVSLRFSDRKLFDDVNIKFTAGNTYGLIGANGAGKSTFLKILAGDIEPSTGHISL 60

Query: 61  GPDERLSVLRQNHFDYEDERVIDVVIMGNETLYSIMKERDAIYMKEDPSDEDGVRAAELE 120
           GPDERLSVLRQNHFDYE+ER IDVVIMGNE LY+IMKERDAIYMK DPS+EDGVRAAELE
Sbjct: 61  GPDERLSVLRQNHFDYEERRAIDVVIMGNEQLYNIMKERDAIYMKADPSEEDCVRAAELE 120

Query: 121 GEPAELGGWEAESEASQLLQNLNISEELHYQNMSELANGDKVKVLLAKALPGKPDVLLLD 180
           G PAELGGWEAESEASQLLQNLNI E+LHYQNMSELANGDKVKVLLAKALPGKPDVLLLD
Sbjct: 121 GIPAELGGWEAESEASQLLQNLNIPEDLHYQNMSELANGDKVKVLLAKALPGKPDVLLLD 180

Query: 181 EPTNGLDIQSITWLEDFLIDFENTVIVVSHDRHFLNKVCTHMADLDFGKIKLFVGNYDFW 240
           EPTNGLDIQSLSWLEDFLIDFENTVIVVSHDRHFLNKVCTHMADLDFGKIKLFVGNYDFW
Sbjct: 181 EPTNGLDIQSLSWLEDFLIDFENTVIVVSHDRHFLNKVCTHMADLDFGKIKLFVGNYDFW 240

Query: 241 KESSELAARLQADRNAKAEEKIKQLQEFVARFSANASKSKQATSRKKMLDKIELEEIVPS 300
           K+SSELAARLQADRNAKAEEKIK+LQEFVARFSANASKSKQATSRKKMLDKIELEEIVPS
Sbjct: 241 KQSSELAARLQADRNAKAEEKIKELQEFVARFSANASKSKQATSRKKMLDKIELEEIVPS 300

Query: 301 SRKYPFVNFKAERKMGNDLLTVENLSVTIDGEKILDNISFILRPGDKTALIGQNDIQTTA 360
           SRKYPF+NFKASREMGNDLLTVENLSVTIDGSKI+DNISFILRPGDK A+IGQNDIQTTA
Sbjct: 301 SRKYPFINFKASREMGNDFLTVSNLSVTIDGEKIIDNISFILRPGDKAAIIGQNDIQTTA 360
```

```
Query: 361 LIRALMGDIEYEGTIKWGVTTSRSYLPKDNSRDFASGRSILEWLRQFASKEEDLNTFLRG 420
           L+RAL   DI+YEGTIKWGVTTSRSYLPKDNS+DFA+ ESILEWLRQFASK EDD+TFLRG
Sbjct: 361 LMRALADDIDYEGTIKWGVTTSRSYLPKDNSKDFATESILEWLRQFASKGEDDDTFLRG 420

Query: 421 FLGRMLFSGDEVNKSVNVLSGGEKVRVMLSKLMLLKSNVLVLDDPTNHLDLESISSLNDG 480
           FLGRMLFSGDEV KSVNVLSGGEKVRVMLSKLMLLKSNVLVLDDPTNHLDLESISSLNDG
Sbjct: 421 FLGRMLFSGDEVKKSVNVLSGGEKVRVMLSKLMLLKSNVLILDDPTNHLDLESISSLNDG 480

Query: 481 LKDFKESIIFASHDHEFIQTLANHIIVLSKNGVIDRIDETYDEFLENTEVQAKVAQLWK 539
           +KDFKES+IFASHDHEFIQT+ANHI+V+SKNGVIDRIDETYDEFL+N EVQA+VA+LWK
Sbjct: 481 IKDFKESVIFASHDHEFIQTIANHIVVISKNGVIDRIDETYDEFLDNPEVQARVARLWK 539
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1645

A DNA sequence (GBSx1740) was identified in *S.agalactiae* <SEQ ID 5091> which encodes the amino acid sequence <SEQ ID 5092>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.71   Transmembrane   14 -  30 (   8 -  35)
   INTEGRAL    Likelihood = -7.70   Transmembrane  384 - 400 ( 382 - 403)
   INTEGRAL    Likelihood = -7.22   Transmembrane  412 - 428 ( 408 - 433)
   INTEGRAL    Likelihood = -5.73   Transmembrane  163 - 179 ( 155 - 180)
   INTEGRAL    Likelihood = -5.52   Transmembrane  322 - 338 ( 320 - 344)
   INTEGRAL    Likelihood = -5.10   Transmembrane  297 - 313 ( 290 - 314)
   INTEGRAL    Likelihood = -4.41   Transmembrane  360 - 376 ( 357 - 377)
   INTEGRAL    Likelihood = -4.35   Transmembrane  438 - 454 ( 437 - 455)
   INTEGRAL    Likelihood = -4.09   Transmembrane  136 - 152 ( 136 - 153)
   INTEGRAL    Likelihood = -3.35   Transmembrane  110 - 126 ( 106 - 128)
   INTEGRAL    Likelihood = -2.28   Transmembrane  232 - 248 ( 232 - 248)
   INTEGRAL    Likelihood = -1.81   Transmembrane  832 - 848 ( 832 - 848)
   INTEGRAL    Likelihood = -1.12   Transmembrane  200 - 216 ( 200 - 216)

----- Final Results -----
                bacterial membrane --- Certainty=0.4885(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC14608 GB:U95840 transmembrane protein Tmp5 [Lactococcus
           lactis]
 Identities = 140/260 (53%), Positives = 182/260 (69%), Gaps = 6/260 (2%)

Query: 16  SFLLPFIIIVCILPTKNIYWGSPTTILASDGFHQVVIPNQALRNILH--GSNSLFVTFTS 73
           SF +P I++V +L    IYWGS  +ILA D +HQVV +   RNILH  GS    VTFTS
Sbjct: 14  SFFIPLILMVIVLAMTGIYWGSSRSILAGDAYHQVVAIHSLYRNILHSGGSGCGFLVTFTS 73

Query: 74  GLGLNFYALSSYYLGSPLSPIVYFPNLKNMPDAIYLLTICKIGLIGLSMFVTLCKRHCKV 133
           GLGLN YA S+YY+GSPL P  +FF++K+MPDA+YL TI K GLIGLS FV+     + K+
Sbjct: 74  GLGLNLYAFSAYYMGSFLMPFTFFFDVKSMPDALYLFTIIKFGLIGLSSFVSFKNMYQKL 133

Query: 134 NKVLLLVISTCYSLMSFSISQTEINMMLDVFILIPLVVLGVDQLLWERKPILYFLSLTAL 193
           + + +L IST ++LMSF  SQ+EI MMLDVFILI PLV+ G+ +L+ ERK  LYF+SL  L
Sbjct: 134 SNLTVLSISTAFALMSFLTSQLEITMMLDVFILLPLIIWGLHRLMDERRRWLYFVSLLIL 193

Query: 194 FIQNYYPGFMTAIFTSLYFIVQITRNTDSKVAFKQPLHFTPLSLLAGMTSSIMILPTYFD 253
           FIQNYYPGFM AIF  LYF   + R T  K ++ + L F   S LAG+ S IM+LP Y D
Sbjct: 194 FIQNYYPGFMVAIPLVLYF---LARMTYEKWSWTXVLDFVVSSTLAGIASLIMLLPMYLD 250

Query: 254 L-TTHGRKLTKVSKMFTENS 272
           L + + + L+ +S +FTENS
```

```
Sbjct: 251 LKSNNSDALSTLSGIPTENS 270
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5093> which encodes the amino acid sequence <SEQ ID 5094>. Analysis of this protein sequence reveals the following:

```
        Possible site: 51
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL   Likelihood = -9.29   Transmembrane   15 -  31 (   6 -  35)
    INTEGRAL   Likelihood = -8.81   Transmembrane  201 - 217 ( 196 - 220)
    INTEGRAL   Likelihood = -6.79   Transmembrane  410 - 426 ( 402 - 428)
    INTEGRAL   Likelihood = -6.05   Transmembrane  230 - 246 ( 227 - 252)
    INTEGRAL   Likelihood = -5.57   Transmembrane  161 - 177 ( 153 - 178)
    INTEGRAL   Likelihood = -4.46   Transmembrane  291 - 307 ( 290 - 311)
    INTEGRAL   Likelihood = -3.82   Transmembrane  133 - 149 ( 130 - 151)
    INTEGRAL   Likelihood = -3.77   Transmembrane  380 - 396 ( 376 - 400)
    INTEGRAL   Likelihood = -3.61   Transmembrane  105 - 121 ( 103 - 124)
    INTEGRAL   Likelihood = -3.45   Transmembrane  832 - 848 ( 830 - 848)
    INTEGRAL   Likelihood = -2.66   Transmembrane  436 - 452 ( 435 - 453)
    INTEGRAL   Likelihood = -2.13   Transmembrane  318 - 334 ( 314 - 336)
    INTEGRAL   Likelihood = -1.54   Transmembrane  356 - 372 ( 355 - 372)
    INTEGRAL   Likelihood = -0.27   Transmembrane   80 -  96 (  80 -  96)

----- Final Results -----
                bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC14608 GB:U95840 transmembrane protein Tmp5 [Lactococcus lactis]
 Identities = 134/269 (49%), Positives = 183/269 (67%), Gaps = 8/269 (2%)

Query: 5   NKWIIAGLASFLPFLSIIFILLSMGIYYNSDKTILASDAFHQYVIFAQNFRNIMH--GS 62
           NKW +  LASF  PL ++ I+L   GIY+ S ++HLA DA+HQYV      +RNI+H  GS
Sbjct: 7   NKWAL--LASFFIPLILMVIVLAMTGIYWGSSRSILAGDAYHQYVAIHSLYRNILHSGGS 64

Query: 63  DSFFYTFTSGLGINFYALMCYYLGSFPSPLLFFPNLTSMPDAIYLFTLIKFGLIGLAACY 122
             F YTFTSGLG+N YA   YY+GSF  P  FFF++ SMPDA+YLFT+IKFGLIGL++
Sbjct: 65  QGFLYTFTSGLGLNLYAFSAYYMGSFLMPFTFFFDVKSMPDALYLFTIIKFGLIGLSSFV 124

Query: 123 SFHRLYPKISAFLMISISVFYSLMSFLTSQMELNSWLDVFILLPLVILGLNKLITENKTR 182
           SF  +Y K+S   ++SIS  ++LMSFLTSQ+E+  WLDVFILLPL+I GL++L+ E K
Sbjct: 125 SFKNMYQKLSNLITVLSISTAFALMSFLTSQLEITMWLDVFILLPLIIWGLHRLMDERKRW 184

Query: 183 TYYLSTSLLFIQNYYFGYMIALFCILYALVCLLRLNDFNNMFIAFVRFTAVSICAALTSA 242
            Y++S+ +LFIQNYYFG+M+A+F +LY L    R+    +   + F   S  A + S
Sbjct: 185 LYFVSLLILFIQNYYFGFMVAIFLVLYFLA---RMTYEKWSWTKVLDPVVSSTLAGIASL 241

Query: 243 LVILPTYLDL-STYGFNLSPIKQLVTNNA 270
           +++LP YLDL S   + LS +  + T N+
Sbjct: 242 IMLLPMYLDLKSNNSDALSTLSGIPTENS 270
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 432/836 (51%), Positives = 569/836 (67%), Gaps = 2/836 (0%)

Query: 16  SFLLPFIIIVCILPTKNIYWGSPTTILASDGFHQYVIFNQALRNILHGSNSLFTTFTSGL 75
           SFL P  II  IL +  IY+ S  TILASD FHQVVIF Q  RNI+HGS+S FTTFTSGL
Sbjct: 14  SFLPFLSIIFILLLSMGIYYNSDKTILASDAFHQYVIFAQNFRNIMHGSDSFFTTFTSGL 73

Query: 76  GLNFYALSSYYLGSFLSPIVVFFNLKNMPDAIYLLTICKIGLIGLSMFVTLCKRHCKVNR 135
           G+NFYAL  YYLGSF SP+++FFNL +MPDAIYL T+ K GLIGL+   + + K++
Sbjct: 74  GINFYALMCYYLGSFFSPLLFFFNLTSMPDAIYLFTLIKFGLIGLAACYSFHRLYPKISA 133

Query: 136 VLLLVISTCYSLMSFSISQIEINMWLDVFLLIPLVVLGVDQLLWERKPILYFLSLTALFI 195
            L++ IS  YSLMSF  SQ+E+N WLDVFLL+PLV+LG+++L+T E K   Y+LS++  LFI
Sbjct: 134 FLMISISVFYSLMSFLTSQMELNSWLDVFILLPLVILGLNKLITENKTRTYYLSISLLFI 193
```

```
Query: 196 QNYYPGPMTAIFTSLYFIVQITRNTDSKVAFKQFLHPTFLSLLAGMTSSIMILPTYFDLT 255
           QNYYPG+M A+F  LY +V + R  D    F  F+ FT +S+ A +TS+++ILPTY DL+
Sbjct: 194 QNYYPGYMIALFCILYAIVCLLRLNDPNKMPIAPVRPTAVSICAALTSALVILPTYLDLS 253

Query: 256 THGEKLTKVSKMFTRNSWYMDLFAKNMIGAYDTTKFGSIPMIYVGLLPLLLSLLYFTIKE 315
           T+GE L+ + ++ T N+W++D+ AK  IG YDTTKF ++PMIYVGL PL+LS++YFT++
Sbjct: 254 TYGENLSPIKQLVTNNAWFLDIPAKLSIGVYDTTKFNALPMIYVGLFPLMLSVIYFTLES 313

Query: 316 VDRRTRLAYGFLIIVIASFYITPLDLFWQGMHAPNMFLHRYSWVLSVLICLLAAECLEY 375
           +D + +LA   L+ F+I SFY+ PLDLFWQGMH+PNMFLHRY+W  S++I LLA E L
Sbjct: 314 IPLKIKLANACLLTPIIISFYLQPLDLFWQGMHSPNMFLHRYAWSFSIVILLACETLSR 373

Query: 376 LDNISWKKILGVNLILVSGPIITFLFKHYHYLNLELLLLLTLFLSAYIILTISFVSKQI 435
           L  ++  K    + L+    + + F + Y++L L L LL++  L  Y I   SF + QI
Sbjct: 374 LKEVTQIKAGFAFIFLIILTSLPYSFSQQYNFLPLTLFLLSVFLLLGYTISLFSFRNSQI 433

Query: 436 PKLVFYPFLIGFVVLEMTLNFTYQLNSLNDEWIFPSRQGYAKYNHSISKLVRKTERNNST 495
           P      F++ F +LE  LNT+YQL  +N EW FPSRQ Y      I+ LV   +N+
Sbjct: 434 PSTFISAFILIFSLLESGLNTYYQLQGINKEWGFPSRQIYNSQLKDINNLVNSVSKKSQP 493

Query: 496 PFKTEWWLGQYTGNDSMKNYNGISQFSSSIRNRSSSQVLDRLGFKSDGTNLNLRYQNNTLI 555
           FFR ER L QTGNDSMK+NY GISQFSS+RNR SS +LDRLGF+S GTNLNLRYQNNT+I
Sbjct: 494 FFRMERLLPQTGNDSMKFNYYGISQFSSVRNRLSSSLLDRLGFQSKGTNLNLRYQNNTII 553

Query: 556 ADSLFGVKYNLTEYPFDKFGFIKKAQDKQTILYKNQFASQLAILTNQVVQDKPFTVNTLD 615
            DSL G+KYNL+E P +KFGF K      T LY+N ++S LAILT  VY+D   VNTLD
Sbjct: 554 MDSLLGIKYNLSEGPPNKFGFIKLNTSGNTTLYQNHYSSPLAILTRNVYKDVNLNVNTLD 613

Query: 616 NQTTLLNQLSGLKETVPEHLLPNSVSGQTTLNKQVFVK-KNKQGNTEITYNITIPKNSQL 674
           NQT LLNQLSG   TYF       +SG    N Q+  +  + Q +  + Y I IPK+SQL
Sbjct: 614 NQTKLLNQLSGKSLTYFNLQPAQLISGANQFNGQISAQASDYQNSVTLNYQINIPKHSQL 673

Query: 675 YVSMPFINFNNEENKIVQISVNNGNDFVPNTLDNAYSFTNIGSFAENSRIKVKPQFPHNDQ 734
           YVS+P I F+N + K ++I  +N  F+  T DNAYSFF++G FA+       F FP N Q
Sbjct: 674 YVSIPNIIPSNPDAKEMRIQTDNHNFI-YTTDNAYSFFDLGYFADAKVATPSPVFPKNKQ 732

Query: 735 VSFPIPHFYGLKLEAYQKAMTVINKRKVKVRTDHNKVIANYTSPNRSSLFFTIPYDRGWK 794
           +SF  PHFY L +E+Y +AM  I ++ V      N VI +Y S  + SL FT+PYD+GW
Sbjct: 733 ISFKEPHFYSLSIESYLEAMNSIKQNVHTYAKSNTVITDYNSKTKGSLIFTLPYDKGWS 792

Query: 795 AYQNNKEIKIPKAQKGFMKINIPKGKGKVTLIFIPYGFKFGVGLSITGIVLFIVYY 850
           A ++ K + + KAQ GF+ + IPKGKG+V L FIP GFK G+ LS  GI+ + + Y
Sbjct: 793 AQKDGKNLPVKKAQQGFLSVTIPKGKGRVILTFIPNGFKLGLSLSCVGIIAYMLLY 848
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1646

A DNA sequence (GBSx1741) was identified in *S.agalactiae* <SEQ ID 5095> which encodes the amino acid sequence <SEQ ID 5096>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4624(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45340 GB:AF000658 ORF1 [Streptococcus pneumoniae]
 Identities = 111/159 (69%), Positives = 136/159 (84%)
```

```
Query: 1    MKLKIITVGKLKEKYLKEGVAEYQKRLNRPSKIETIELADEKTPDKASISENQRILDIEG 60
            MK+K++TVGKLKEKYLK+G+AEY KR++RP+K E IEL+DEKTPDKAS SENQ+IL+IEG
Sbjct: 1    MKIKVVTVGKLKEKYLKDGIAEYSKRISRPAKPEMIELSDEKTPDKASSENQKILEIEG 60

Query: 61   ERILSKIGERDYVIGLAIEGKQLPSESPSHLIDQKMISGYSTITFVIGGSLGLSQKVKKR 120
            +RILSKI +RD+VI LAIEGK   SE FS  +++   I G+ST+TF+IGGSLGLS   VK R
Sbjct: 61   QRILSKIADRDFVIVLAIEGKTFFSEEFSKQLEETSIKGFSTLTFIIGGSLGLSSSVKNR 120

Query: 121  ADYLMSPGLLTLPHQLMKLVLMEQIYRAFMIRQGTPYHK 159
             A+  +SPG LTLPHQLM+LVL+EQIYRAF I+QG PYHK
Sbjct: 121  ANLSVSPGRLTLPHQLMRLVLVEQIYRAFTIQQGFPYHK 159
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5097> which encodes the amino acid sequence <SEQ ID 5098>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4462(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 112/159 (70%), Positives = 133/159 (83%)

Query: 1    MKLKIITVGKLKEKYLKEGVAEYQKRLNRPSKIETIELADEKTPDKASISENQRILDIEG 60
            MK+K+I VGKLKE+YLK+G++EYQKRL+RP + E IEL DE+TPDKAS ++NQ I+  E
Sbjct: 1    MKVKLICVGKLKERYLKDGISEYQKRLSRPCQPEMIELITDERTPDKASFADNQLIMSKKA 60

Query: 61   ERILSKIGERDYVIGLAIEGKQLPSESPSHLIDQKMISGYSTITFVIGGSLGLSQKVKKR 120
            +RI  KIGERD+VI LAIEGKQ PSE+PS LI    + GYSTITF+IGGSLGL    +KKR
Sbjct: 61   QRIHKKIGERDFVIALAIEGKQFPSETFSELISGVTVKGYSTITFIIGGSLGLDSIIKKR 120

Query: 121  ADYLMSPGLLTLPHQLMKLVLMEQIYRAFMIRQGTPYHK 159
             A+ LMSPGLLTLPHQLM+LVL EQIYRAFMI QG+PYHK
Sbjct: 121  ANMLMSPGLLTLPHQLMRLVLTEQIYRAFMITQGSPYHK 159
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1647

A DNA sequence (GBSx1742) was identified in *S.agalactiae* <SEQ ID 5099> which encodes the amino acid sequence <SEQ ID 5100>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3785(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1648

A DNA sequence (GBSx1743) was identified in *S.agalactiae* <SEQ ID 5101> which encodes the amino acid sequence <SEQ ID 5102>. This protein is predicted to be a serine protease. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4533(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9445> which encodes amino acid sequence <SEQ ID 9446> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45334 GB:AF000658 putative serine protease [Streptococcus pneumoniae]
 Identities = 215/370 (58%), Positives = 278/370 (75%), Gaps = 20/370 (5%)

Query: 4   NDNIPNGGVTKTSKVNTNNITPTTYAVKKVQNSVVSVINYKQQESRSDLSDFYSHPPGNQ 63
            N++  N  +T+T+   Y N   TT+AV KV++++VVSVI Y    S        PGN
Sbjct: 46  NNSNNNSTITQTA---YKNENSTTQAVNKVKDAVVSVITYSANRQNS--------VPGND 94

Query: 64  GGNTDKGLQVYGEGSGVIYKKDGKNAYVVTNNHVIDGAKQIEIQLADGSKAVGKLWGSDT 123
              +TD   ++  EGSGVIYKK+ K AY+VTNNHVI+GA +++I+L+DG+K  G++VG+DT
Sbjct: 95  DTDTDSQ-RISSEGSGVIYKKNDKEAYIVTNNHVINGASKVDIRLSDGTKVPGEIVGADT 153

Query: 124 YSDLAVVKIPSDKVSNIAEPADSSKLNIGETAIAIGSPLGTEYANSVTQGIVSSLKRTVT 183
            +SD+AVVKI S+KV+ +AEP DSSKL +GETAIAIGSPLG+EYAN+VTQGIVSSL R V+
Sbjct: 154 PSDIAVVKISSEKVTTVAEPGDSSKLTVGETAIAIGSPLGSEYANTVTQGIVSSLNRNVS 213

Query: 184 MTNERGQTVSTNAIQTDAAINPGNSGGALINIEGQVIGINSSKISSTSNQTSGQSSGNSV 243
            + +E+GQ +ST AIQTD AINPGNSGG LINI+GQVIGI SSKI++          + G SV
Sbjct: 214 LKSEDGQAISTKAIQTDTAINPGNSGGPLINIQGQVIGITSSKIAT--------NGGTSV 265

Query: 244 EGMGFAIPSNDVVKIINQLESNGQVERPALGISMAGLSNLPSDVISKLKIPSNVTNGIVV 303
            EG+GFAIP+ND + II QLE NG+V RPALGI M  LSN+ +  I +L IPSNVT+G++V
Sbjct: 266 EGLGFAIPANDAINIIEQLEKNGKVTRPALGIQMVNLSNVSTSDIRRLNIPSNVTSGVIV 325

Query: 304 ASIQSGMPAQGKLKKYDVITKVDDKEVVSPSDLQSLLYGHQVGDSITVTFYRGENKQTVT 363
             S+QS MPA G L+KYDVITKVDDKE+ S +DLQS LY H +GD+I +T+YR   ++T +
Sbjct: 326 RSVQSNMPANGHLEKYDVITKVDDKEIASSTDLQSALYNHSIGDTIKITTYYRNGKEETTS 385

Query: 364 IKLTKTSKDL 373
            IKL K+S DL
Sbjct: 386 IKLNKSSGDL 395
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5103> which encodes the amino acid sequence <SEQ ID 5104>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood = -8.76   Transmembrane    11 -  27 (   6 -  31)

----- Final Results -----
             bacterial membrane --- Certainty=0.4503(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 250/375 (66%), Positives = 299/375 (79%), Gaps = 5/375 (1%)
```

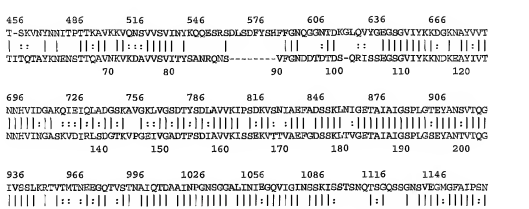
```
Query: 3   HNDNIPNGGVTKTSKVNYNNITPTTKAVKKVQNSVVSVINYKQQRSRSDLSDPYSHFPGN 62
           H+ +  N G   TS + +NN T TTKAVK VQN+VVSVINY+    S S LS+ Y+  FG
Sbjct: 34  HSPSKINSGKATTSNMVFNNTTNTTKAVKAVQNAVVSVINYQDNPS-SSLSNPYTKLFGE 92

Query: 63  QGG--NTDKGLQVYGEGSGVIYKKDGKNAYVVTNNHVIDGAKQIEIQLADGSKAVGKLVG 120
                N D  L ++ EGSGVIY+KDG +AYVVTNNHVIDGAK+IEI +ADGSK VG+LVG
Sbjct: 93  GRSKENKDAELSIFSEGSGVIYRKLGNSAYVVTNNHVIDGAKRIKILMADGSKVVGELVG 152

Query: 121 SDTYSDLAVVKIPSDKVSNIARFADSSKLNIGETAIAIGSPLGTEYANSVTQGIVSSLKR 180
            +DTYSDLAVVKI SDK+  +AEFADS+KLN+GE AIAIGSPLGT+YANSVTQGIVSSL R
Sbjct: 153 ADTYSDLAVVKISSDKIKTVAEFADSTKINVGEVAIAIGSPLGTQYANSVTQGIVSSLSR 212

Query: 181 TVTMTNEGQVTVSTNAIQTDAAINPGNSGGALINIEGQVIGINSSKISSTSNQTSGCSSG 240
            TVT+ NE G+TVSTNAIQTDAAINPGNSGG LINIEGQVIGINSSKISST    ++G S
Sbjct: 213 TVTLKNENGETVSTNAIQTDAAINPGNSGGPLINIEGQVIGINSSKISSTPTGSNGNS-- 270

Query: 241 NSVEGMGFAIPSNDVVKIINQLESNGQVERPALGISMAGLSNLPSDVISKLKIPSNVTNG 300
             +VEG+GFAIPS DV+KII QLE+NG+V RPALGISM  L++L ++ +S++ IP++VT G
Sbjct: 271 GAVEGIGFAIPSTDVIKIIKQLETNGEVIRPALGISMVNLNDLSTNALSQINIPTSVTGG 330

Query: 301 IVVASIQSGNPAQGKLKKVDVITKVDDKEVVSPSDLQSLLYGHQVGDSITVTFYRGENKQ 360
            IVVA ++ GNPA GKL +YDVIT+D K V S SDLQS LYGH + D+I VTFYRG  K+
Sbjct: 331 IVVASVKEGNPASGKLAQYDVITEIDGKTVNSISDLQSSLYGHDINDTIKVTFYRGPTKK 390

Query: 361 TVTIKLTKTSKDLAK 375
              IKLTKT++DL K
Sbjct: 391 KADIKLTKTTQDLTK 405
```

A related GBS gene <SEQ ID 8873> and protein <SEQ ID 8874> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    12.68
GvH: Signal Score (-7.5): -1.33
     Possible site: 21
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   4.56 threshold:  0.0
   PERIPHERAL  Likelihood =  4.56     301
 modified ALOM score:  -1.41

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
57.4/75.6% over 386aa
                                                                  Streptococcus
pneumoniae
  GP|2109443| putative serine protease Insert characterized

ORF02135(307 - 1506 of 1827)
GP|2109443|gb|AAC45334.1||AF000658(9 - 395 of 397) putative serine protease (Streptococcus
pneumoniae)
%Match = 34.6
%Identity = 57.3  %Similarity = 75.6
Matches = 223  Mismatches = 89  Conservative Sub.s = 71

228       258       288       318       348       378       399       429
RLSTSCGYFLFLAFKV*LRSLS*D*YKNLRR*LPVKKKLVSSLLKCSLIIIVSFAGGAFASFVMNH---MDNIPNGGVTK
                                :   :   |  :    ::  ::|::||  ||: || : :     :: |
                               MEANMKHLKTFYKKWFQLLVVIVISFFSGALGSFSITQLTQKSSVNNSNNNS
                                        10        20        30        40        50
```



Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1649

A DNA sequence (GBSx1744) was identified in *S.agalactiae* <SEQ ID 5105> which encodes the amino acid sequence <SEQ ID 5106>. This protein is predicted to be SPSpoJ (spo0J). Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4152(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45335 GB:AF000658 SPSpoJ [Streptococcus pneumoniae]
 Identities = 138/257 (53%), Positives = 188/257 (72%), Gaps = 5/257 (1%)

Query: 1   MEYLETTINNHIAPNPYQPRLEPNTKELEELANSIKINGLIQPIIVRPSAVPGYELVAGE 60
            ME  E I+I  I  NPYQPR EF+ ++L+ELA SIK NG+IQPIIVR S V GYE++AGE
Sbjct: 1   MEKFEMISITDIQKNPYQPRKEFDREKLDELAQSIKENGVIQPIIVRQSPVIGYEILAGE 60

Query: 61  RRLRAAKLAKLESIPAIIKSYNNDDSMQLAIVENLQRSNLSPIEEAKAYSQLLQKKSMTH 120
            RR KA+ LA L SIPA++K  ++ + M  +I+ENLQR NL+PIEKA+AY  L++ K  TH
Sbjct: 61  RRYRASLLAGLRSIPAVVKQISDQEMMVQSIIENLQRENLNPIEERARAYVSLVE-KGFTH 119

Query: 121 EELAKYMGKSRPYISNTIRLLNLPPLITSAIEEGKLSSGHARALLSLPDASQQEDWYQRI 180
             E+A   GKSRPYISN+IRLL+LP  I S +S GKLS  HAR+L+ L +  QQ  ++QRI
Sbjct: 120 AEIADKEGKSRPYISNSIRLLSLPEQILSEVESNGKLSQAHARSLVGL-NKEQQDYFPQRI 178
```

```
Query: 181 LTEDISVRRLEKLLKQEKKTNHKSLQNKDVFLKHQRNELAQFLGSKVKLTINKDGAGNIK 240
           + EDISVR+LE LL ++K+   K  Q  + F++++E +L + LG  V++ ++K  +G I
Sbjct: 179 IEEDISVRKLEALLTEKKQ---KKQQKTNHFIQNEKQLRKLLGLDVEIKLSKKDSGKII 235

Query: 241 IAFANQEELNRIINTLK 257
           I+F+NQEE +RIIN+LK
Sbjct: 236 ISFSNQEEYSRIINSLK 252
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5107> which encodes the amino acid sequence <SEQ ID 5108>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1758(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 146/256 (57%), Positives = 191/256 (74%), Gaps = 1/256 (0%)

Query: 2   EYLETININHIAPMPYQPRLEFNTKELEELANSIKINGLIQPIIVRPSAVFGYELVAGER 61
           E L  + I  I  NPYQPR++FN +EL++LA SIK NGLIQPIIVR S +FGYELVAGER
Sbjct: 14  ELLIDLPIEDIVTNPYQPRIQFNQRELQDLATSIKSNGLIQPIIVRKSDIFGYELVAGER 73

Query: 62  RLRAAKLAKLESIPAIIKSYNNDDSMQLAIVENLQRSNLSPIEEAKAYSQLLQKKSMTHE 121
           RL+A+K+A L+ +PAIIK  +  +SMQ AIVENLQRSNL+ IEEAKAY  L++KK MTH+
Sbjct: 74  RLKASKMAGLKKVPAIIKKISTLESMQQAIVENLQRSNLNAIEEAKAYQLLVEKKHMTHD 133

Query: 122 ELAKYMGKSRPYISNYIRLLNLPPLITSAIEEGKLSSGHARALLSLPDASQQKDWYQRIL 181
           E+AKYMGKSRPYISNYIRLL LP  I  AIEEGK+S+GHARALL+L D  QQ      +I
Sbjct: 134 ELAKYMGKSRPYISNYLRLLQLPAPIIKAIEEGKISAGHARALLTLSDDKQQLYLTHKIQ 193

Query: 182 TEDISVRRLEKLLKQEKKTNHKSLQNKDVFLKHQENELAQFLGSKVKLTINKDGAGNIKI 241
            E +SVR++E+L+      ++  S + K++F    E +LA+ LG  V + +  + +G ++I
Sbjct: 194 NEGLSVRQIEQLV-TSTPSSKLSKKTKNIFATSLEKQLAKSLGLSVNMKLTANESGYLQI 252

Query: 242 AFANQEELNRIINTLK 257
           +F+N +ELNRIIN LK
Sbjct: 253 SFSNDDELNRIINKLK 268
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1650

A DNA sequence (GBSx1745) was identified in *S.agalactiae* <SEQ ID 5109> which encodes the amino acid sequence <SEQ ID 5110>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane     2 -  18 (   1 -  18)

----- Final Results -----
                bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10297> which encodes amino acid sequence <SEQ ID 10298> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5111> which encodes the amino acid sequence <SEQ ID 5112>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3646(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 353/455 (77%), Positives = 401/455 (87%), Gaps = 6/455 (1%)

Query: 32  MTENEQLFWNRVLRLSRSQIAPAAYEFFVLEARLLKIEHQTAVITLDNIEMKKLPWEQNL 91
           MTENEQ+FWNRVLEL++SQ+  A YEFFV +ARLLK++    A I LD  +ME+LPWE+NL
Sbjct: 1   MTENEQIFWNRVLELAQSQLKQATYEFFVHDARLLKRVDKHIATIYLGD--QMKELPWEKNL 58

Query: 92  GPVILTAGFEIPNAEITANYV-SNDLHLQETSFS-NVQQSSNEVNTLPIRKIDSNLKEKY 149
             VILTAGFE++NA+I+ +YV   DL +++      N +      +N+LP   + S+L  KY
Sbjct: 59  KDVILTAGFEVINAQISVDYVFEEDLMIEQNQTKINQKPKQQALNSLPT--VTSDLNSKY 116

Query: 150 TFANFVQGDENRWAVSASIAVADSPGTTYNPLFIWGGPGLGKTHLLNAIGNQVLRDNPNA 209
           +F NF+QGDENRWAV+ASIAVA++PGTTYNPLFIWGGPGLGKTHLLNAIGN VL +NPNA
Sbjct: 117 SFENFIQGDENRWAVAASIAVANTPGTTYNPLFIWGGPGLGKTHLLNAIGNSVLLENPNA 176

Query: 210 RVLYITAENFINEFVSHIRLDSMEELKEKFRNLDLLLIDDIQSLAKKTLGGTQEEFFNTF 269
           R+ YITAENFINEFV HIRLD+M+ELKEKFRNLDLLLIDDIQSLAKKTL GTQEEFFNTF
Sbjct: 177 RIKYITAENFINEFVIHIRLDTMDELKEKFRNLDLLLIDDIQSLAKKTLSGTQEEFFNTF 236

Query: 270 NALHTNDKQIVLTSDRNPNQLNDLEERLVTRFSWGLPVNITPPDFETRVAILTNKIQEYP 329
           NALH N+KQIVLTSDR P+ LNDLE+RLVTRF WGL VNITPPDFETRVAILTNKIQEY
Sbjct: 237 NALHNDNKQIVLTSDRTPDHLNDLEDRLVTRFKWGLTVNITPPDFETRVAILTNKIQEYN 296

Query: 330 YDFPQDTIEYLAGEFDSNVRELEGALKNISLVADFKHAKTITVDIAAEAIRARKNDGPIV 389
           + FPQDTIEYLAG+FDSNVR+LEGALK+ISLVA+FK   TITVDIAAEAIRARK DGP +
Sbjct: 297 FIFPQDTIEYLAGQFDSNVRDLEGALKDISLVANFKQIDTITVDIAAEAIRARKQDGPKM 356

Query: 390 TVIPIEEIQIQVGKFYGVTVKEIKATKRTQDIVLARQVAMYLARRMTDNSLPKIGKEFGG 449
           TVIPIEEIQ QVGKFYGVTVKEIKATKRTQ+IVLARQVAM+LARRMTDNSLPKIGKEFGG
Sbjct: 357 TVIPIEEIQAQVGKFYGVTVKEIKATKRTQNIVLARQVAMFLARRMTDNSLPKIGKEFGG 416

Query: 450 RDHSTVLHAYNKIKNMVAQDDNLRIEIETIKNKIR 484
           RDHSTVLHAYNKIKNM++QD++LRIEIETIKNKI+
Sbjct: 417 RDHSTVLHAYNKIKNMISQDESLRIEIETIKNKIK 451
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1651

A DNA sequence (GBSx1746) was identified in *S.agalactiae* <SEQ ID 5113> which encodes the amino acid sequence <SEQ ID 5114>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0556(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45337 GB:AF000658 beta subunit of DNA polymerase III
          [Streptococcus pneumoniae]
 Identities = 278/378 (73%), Positives = 324/378 (85%)

Query: 1   MIHFSINKNFFLHALTVTKRAISHKNAIPILSTVKIEVTRDAIILTGSNGQISIENTIPA 60
           MIHFSINKN FL AL +TKRAIS KNAIPILSTVKI+VT + + L GSNGQISIEN I
Sbjct: 1   MIHFSINKNLFLQALNTTKRAISSKNAIPILSTVKIDVTNEGVTLIGSNGQISIENFISQ 60

Query: 61  SNENAGLLVTNPGSILLEAGFFINIISSLPDVTLEPTSIEQHQIVLTSGKSEITLKGKDV 120
           NE+AGLL+T+ GSILLEA FFIN++SSLPDVTL+P EIEQ+QIVLTSGKSEITLKGKD
Sbjct: 61  KNEDAGLLITSLGSILLEASFFINVVSSLPDVTLDFKEIEQNQIVLTSGKSEITLKGKDS 120

Query: 121 DQYPRLQEMTTDTPLLETKLLKSIINETAPAASQQSRPILTGVHLVISQNKYFKAVAT 180
           +QYPR+QE++  TPL LETKLLK IINETAPAAS QESRPILTGVH V+SQ+K  K VAT
Sbjct: 121 EQYPRIQEISASTPLLETKLLKKIINETAPAASTQESRPILTGVHFVLSQHKELKTVAT 180

Query: 181 DSHRMSQRTFQLEKSANNPDLVVPSKSLREPSAVFTDDIETVEVFFSDSQMLFRSENISF 240
           DSHR+SQ+   LEK++++PD+V+PS+SLREPSAVFTDDIETVE+FF+++Q+LFRSENISF
Sbjct: 181 DSHRLSQKKLTLEKNSDDFDVVIPSRSLREPSAVFTDDIETVEIFFANNQILFRSENISF 240

Query: 241 YTRLLEGNYPDTDRLLTNQFETEIIFNTNALRHAMERAYLISNATQNGTVRLEIQNETVS 300
           YTRLLEGNYPDTDRL+   F T I FN   LR +MERA L+S+ATQNGTV+LEI++  VS
Sbjct: 241 YTRLLEGNYPDTDRLIPTDFNTTITFNVVNLRQSMERARLLSSATQNGTVKLEIKDGVVS 300

Query: 301 AEVNSPEVGKVNEELDTVSLKGDSLNISFNPTYLISSLKAVKSETVTIRFISPVRPTTLT 360
           AEV+SPEVGKVNEE+DT  + G+ L ISFNPTYLI+SLKA+ SE VTI FIS VRPFTL
Sbjct: 301 AEVHSPEVGKVNEEIDTQVTGEDLTISFNPTYLIDSLKALNSEKVTISFISAVRPFTLV 360

Query: 361 PGEDTEDFIQLITPVRTN 378
           P +  EDF+QLITPVRTN
Sbjct: 361 PADTDEDFMQLITPVRTN 378
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5115> which encodes the amino acid sequence <SEQ ID 5116>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.70   Transmembrane   67 -  83 (  67 -  83)

----- Final Results -----
                bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 295/378 (78%), Positives = 334/378 (88%)

Query: 1   MIHFSINKNFFLHALTVTKRAISHKNAIPILSTVKIEVTRDAIILTGSNGQISIENTIPA 60
           MI FSIN+  F+HAL  TKRAIS KNAIPILS++KIEVT   + LTGSNGQISIENTIP
Sbjct: 1   MIQFSINRTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSNGQISIENTIPV 60

Query: 61  SNENAGLLVTNPGSILLEAGFPFINIISSLPDVTLEPTSIEQHQIVLTSGKSEITLKGKDV 120
           SNENAGLL+T+PG+ILLEA FFINIISSLPDISINVKSIEQHQVVLTSGKSEITLKGKDV 120

Query: 121 DQYPRLQEWTTDTPLTLETKLLKSIINETAPAASQQSRPILTGVHLVISQNKYFKAVAT 180
           DQYPRLQE++T+ PL L+TKLLKSII ETAPAAS QESRPILTGVH+V+S +K FKAVAT
Sbjct: 121 DQYPRLQEVSTENPLLLKTKLLKSIIAETAPAASLQESRPILTGVHIVLSNHKDFKAVAT 180

Query: 181 DSHRMSQRTPQLEKSANNFDLVVPSKSLREPSAVFTDDIETVEVFFSDSQMLFRSENISF 240
           DSHRMSQR   L+ ++ +FD+V+PSKSLREPSAVFTDDIETVEVFFS SQ+LFRSE+ISF
Sbjct: 181 DSHRMSQRLITLDNTSADFDVVIPSKSLREPSAVFTDDIETVEVFFSPSQILFRSEHISF 240

Query: 241 YTRLLEGNYPDTDRLLTNQFETEIIFNTNALRHAMERAYLISNATQNGTVKLEITQNHIS 300
           YTRLLEGNYPDTDRLL  +FSTE++FNT  +LRHAMERA LISNATQNGTV+LEI    +S
Sbjct: 241 YTRLLEGNYPDTDRLLMTEFSTEVVFNTQSLRHAMERAFLISNATQNGTVKLEITQNHIS 300
```

```
Query: 301 AHVNSPEVGKVNEKLDTVSLKGDSLNISFNPTYLIESLKAVKSETVTIRFISPVRPFTLT 360
           AHVNSPEVGKVNE+LD VS  G  L ISFNPTYLIESLKA+KSETV I F+SPVRPFTLT
Sbjct: 301 AHVNSPEVGKVNEDLDIVSQSGSDLTISFNPTYLIESLKAIKSETVKIHFLSPVRPFTLT 360

Query: 361 PGEDTEDFIQLITPVRTN 378
           PG++ E FIQLITPVRTN
Sbjct: 361 PGDEEESFIQLITPVRTN 378
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1652

A DNA sequence (GBSx1747) was identified in *S.agalactiae* <SEQ ID 5117> which encodes the amino acid sequence <SEQ ID 5118>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0857(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10299> which encodes amino acid sequence <SEQ ID 10300> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC00282 GB:AF008220 YtlR [Bacillus subtilis]
 Identities = 83/298 (27%), Positives = 138/298 (45%), Gaps = 35/298 (11%)

Query: 19  YIIANPHAGNKNASTIVGKIQE--LYHTEDISVFYTEQKDDEK--KQVINILRSFKESDH 74
           + I NP AG++N    +   IQ+  +    +   F TE      +  I+ ++ +K
Sbjct: 5   FFIINPTAGHRNGLRVWKSIQKELIKRKVEHRSFLTEHPGHAEVLARQISTIQEYKLK-R 63

Query: 75  LMIIGGDGTLSKVMTYLPQ--HIPCTYYPVGSGNDFARALKIPNL---------KSTLTA 123
           L++IGGDGT+ +V+  L     I  ++ P G+ NDF+R   I  +          K  LT
Sbjct: 64  LIVIGGDGTMHEVVNGLKDVDDIRLSFVPAGAYNDFSRQFSIKKIDLIQEIKKVKRPLT- 122

Query: 124 IQTERLKEINCFIYDKGLIL---NSLDLGFAAYVVWKASNSKIKNILNRYRLGKITYIVI 180
            +T  L  +N F+ DK  IL    N + +GF AIV  KA      ++ +    RL  + Y +
Sbjct: 123 -RTFHLGSVN-FLQDKSQILYFMNHIGIGFDAYVNKKAMEFPLRRVFLFLRLRPLVYPL- 179

Query: 181 AIKSLLHSSK------VQVLVEGETQQQIKLNDLYFFALANNTYFGGGITIWPKASALTA 234
            S LH+S       +    E ET +    +D++F  ++N+ ++GGG+   P A+
Sbjct: 180 ---SHLWASATFKPFTLACTTEDETRE---FHDVWEAVVSNHPFYGGGMKAAPLANPREK 233

Query: 235 ELDMVYAKGHTFLKRLSILLSLVFKRHTTSKSIKHQTFKAMTVYFPKNSLIEIDGEIV 292
               D+V  +  FLK+  +L  + F +HT     +  K +T Y       DGEI+
Sbjct: 234 TFDIVIVKNQPFLKKYWLLCLMAFGKHTKMDGVIMFKAKDITFYTKDKIPFHADGEIM 291
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1653

A DNA sequence (GBSx1748) was identified in *S.agalactiae* <SEQ ID 5121> which encodes the amino acid sequence <SEQ ID 5122>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
                bacterial cytoplasm --- Certainty=0.3792(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45338 GB:AF000658 ORFX [Streptococcus pneumoniae]
 Identities = 46/63 (73%), Positives = 57/63 (90%)

Query: 1  MYQVGSLVEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK 60
          MYQVG+ VEMKKPHAC IK TGKKAN+W++ RVGADIKI+C+NC+HV+MM RYDFERK+
Sbjct: 1  MYQVGNFVEMKKPHACTIKSTGKKANRWEITRVGADIKIKCSNCEHVVMMGRYDFERKWN 60

Query: 61 KVL 63
           K++
Sbjct: 61 KII 63
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5123> which encodes the amino acid sequence <SEQ ID 5124>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.4038(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 63/65 (96%), Positives = 64/65 (97%)

Query: 1  MYQVGSLVEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK 60
          MYQ+GS VEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK 60
Sbjct: 1  MYQIGSFVEMKKPHACVIKETGKKANQWKVLRVGADIKIQCTNCQHVIMMSRYDFERKLK 60

Query: 61 KVLQP 65
           KVLQP
Sbjct: 61 KVLQP 65
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1654

A DNA sequence (GBSx1749) was identified in *S.agalactiae* <SEQ ID 5125> which encodes the amino acid sequence <SEQ ID 5126>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.99   Transmembrane   48 -  64 (  47 -  66)

----- Final Results -----
                bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1655

A DNA sequence (GBSx1750) was identified in *S.agalactiae* <SEQ ID 5127> which encodes the amino acid sequence <SEQ ID 5128>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4171(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1656

A DNA sequence (GBSx1751) was identified in *S.agalactiae* <SEQ ID 5129> which encodes the amino acid sequence <SEQ ID 5130>. This protein is predicted to be GTP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3952(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8875> which encodes amino acid sequence <SEQ ID 8876> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:      0.53
GvH: Signal Score (-7.5): -0.13
     Possible site: 29
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   1.48 threshold:  0.0
   PERIPHERAL  Likelihood =  1.48     195
 modified ALOM score:  -0.60

*** Reasoning Step: 3

----- Final Results -----
              bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07770 GB:AP001520 GTP-binding protein [Bacillus halodurans]
 Identities = 223/329 (67%), Positives = 273/329 (82%), Gaps = 5/329 (1%)

Query: 1   MVEVPDERLQKLTELITPKKTVPTTPETDIAGIVKGASKGEGLGNKFLANIREVDAIVH 60
```

```
              +VEVPD RLQKLTEL+ PKKTVPT FEPTDIAGIV+GASKGEGLGN+FL++IR+VDAI H
Sbjct: 43    IVEVPDPRLQKLTELVNPKKTVPTAFEPTDIAGIVEGASKGEGLGNQFLSHIRQVDAISH 102

Query: 61    VVRAFDDENVMREQGREDAFVDPIADIDTINLELILADLESSINKRYARVEKMARTQKDKE 120
              VVR FDDEN+     G     VDPI DI  INLELILADLES++KR++RV+K+A+T KDKE
Sbjct: 103   VVRCFDDENITHVSGS----VDPIRDISVINLELILADLESVDKRPSRVQKLAKT-KDKE 157

Query: 121   SVAEFNVLQKIKPVLEDGKSARTIEFTEEEAKVVKGLFLLITKPVLYVANVDEDKVADPD 180
              +VAE  VL+K+K   E+ K AR+IEFTEE+ K+VKGL LLT+KPVLYVANV ED V  PD
Sbjct: 158   AVAELEVLEKLKDAFENEKPARSIEFTEEQQKIVKGLHLLTSKPVLYVANVSEDDVLSPD 217

Query: 181   DIDYVNQIRAFARTGNAEVVVISARAEEEISELDDEDKLEPLEAICLTESGVDKLTRAAY 240
              D  +V +++AFA  EN+EV+V+SA+ EEEI+ELD E+K  FLE +G+ ESG+D+L RAAY
Sbjct: 218   DNPFVQKVKAFAAENSEVIVVSAKIEEEIAELDGEEKAMFLEELGIQESGLDQLIRAAY 277

Query: 241   HLLGLGTYFTAGKEVRAWTPKRGIKAPQAASIIHSDFERGFIRAVTMSYDDLIQYGSEK 300
               LLGL TYFTAGE+EVRAWTF++G KAPQAA IIHSDFE+GFIRA T+SY+DL++ GS
Sbjct: 278   SLLGLQTYFTAGEQEVRAWTFRKGTKAPQAAGIIHSDFEKGFIRAETVSYNDLVEAGSMA 337

Query: 301   AVKEAGRLREEGKSYIVQDGDIMEFRFNV 329
               KE G+ +R EGKSY+VQDGD++ FRFNV
Sbjct: 338   VAKERGKVRLEGKEYVVQDGDVIHFRFNV 366
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5131> which encodes the amino acid sequence <SEQ ID 5132>. Analysis of this protein sequence reveals the following:

```
        Possible site: 29

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB07770 GB:AP001520 GTP-binding protein [Bacillus halodurans]
 Identities = 259/371 (69%), Positives = 314/371 (83%), Gaps = 5/371 (1%)

Query: 1     MALTAGIVGLPNVGKSTLFNAITKAGAEAANYPFATIDPNVGMVEVPDERLQKLTELITP 60
              MALT GIVGLPNVGKSTLFNAIT+AGAE+ANYPF TIDPNVG+VEVPD RLQKLTEL+ P
Sbjct: 1     MALTTGIVGLPNVGKSTLFNAITQAGAESANYPFCTIDPNVGIVEVPDPRLQKLTELVNP 60

Query: 61    KKTVPTFFETDIAGIVKGASRGEGLGNKFLANIRSIDAIVHVVRAFDDENVMREQGRED 120
              KKTVPT FEPTDIAGIV+GAS+GEGLGN+FL++IR++DAI HVVR FDDEN+     G
Sbjct: 61    KKTVPTAFEPTDIAGIVEGASKGEGLGNQFLSHIRQVDAISHVVRCFDDENITHVSGS-- 118

Query: 121   AFVDPIADIDTINLELILADLESSINKRYARVEKMARTQKDKESVAEFNVLQKIKPVLEDG 180
               VDPI DI  INLELILADLES++KR++RV+K+A+T KDKE+VAE  VL+K+K   E+
Sbjct: 119   --VDPIRDISVINLELILADLESVDKRFSRVQKLAKT-KDKEAVAELSVLEKLKDAFENE 175

Query: 181   KSARTIEFTEDEAKVVKGLFLLITKPVLYVANVDEDKVANPDGIDYVKQIRDFAATENAE 240
              K AR+IEFTE++ K+VKGL LLT+KPVLYVANV ED V +PD   +V++++ FAA EN+E
Sbjct: 176   KPARSIEFTEEQQKIVKGLHLLTSKPVLYVANVSEDDVLSPDDNPFVQKVKAFAAESNSE 235

Query: 241   VVVISARAEEEISELDDEDKEEFLEAIGLTESGVDKLTRAAYHLLGLGTYFTAGEKEVRA 300
              V+V+SA+ EEEI+ELD E+K  FLE +G+ ESG+D+L RAAY LLGL TYFTAGE+EVRA
Sbjct: 236   VIVVSAKIEEEIAELDGEEKAMFLEELGIQESGLDQLIRAAYSLLGLQTYFTAGEQEVRA 295

Query: 301   WTFKRGIKAPQAAGIIHSDFERGFIRAVTMSYDDLMTYGSEKAVKEAGRLREEGKEYVVQ 360
              WTF++G KAPQAAGIIHSDFE+GFIRA T+SY+DL+  GS     KE G++R EGKEYVVQ
Sbjct: 296   WTFRKGTKAPQAAGIIHSDFEKGFIRAETVSYNDLVEAGSNAVAKERGKVRLEGKEYVVQ 355

Query: 361   DGDIMEFRFNV 371
              DGD++ FRFNV
Sbjct: 356   DGDVIHFRFNV 366
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 316/329 (96%), Positives = 322/329 (97%)

Query: 1   MVEVPDERLQKLTELITPKKTVPTTPEFTDIAGIVKGASKGEGLGNKFLANIREVDAIVH 60
            MVEVPDERLQKLTELITPKKTVPTTPEFTDIAGIVKGAS+GEGLGNKFLANIRE+DAIVH
Sbjct: 43  MVEVPDERLQKLTELITPKKTVPTTPEFTDIAGIVKGASRGEGLGNKFLANIREIDAIVH 102

Query: 61  VVRAFDDENVMREQGREDAFVDPIADIDTINLELIADLESINKRYARVEKMARTQKDKE 120
            VVRAFDDENVMREQGREDAFVDPIADIDTINLELIADLESINKRYARVEKMARTQKDKE
Sbjct: 103 VVRAFDDENVMREQGREDAFVDPIADIDTINLELIADLESINKRYARVEKMARTQKDKE 162

Query: 121 SVAEFNVLQKIKPVLEDGKSARTIEFTKEEAKVVKGLFLLTTKPVLYVANVDEDKVADPD 180
            SVAEFNVLQKIKPVLEDGKSARTIEFTE+EAKVVKGLFLLTTKPVLYVANVDEDKVA+PD
Sbjct: 163 SVAEFNVLQKIKPVLEDGKSARTIEFTEDEAKVVKGLFLLTTKPVLYVANVDEDKVANPD 222

Query: 181 DIDYVNQIRAFAETENAEVVVISARAEEEISELDDEDKLEFLEAIGLTESGVDKLTRAAY 240
             IDYV QIR FA TENAEVVVISARAEEEISELDDEDK EFLEAIGLTESGVDKLTRAAY
Sbjct: 223 GIDYVKQIRDFAATENAEVVVISARAEEEISELDDEDKEEFLEAIGLTESGVDKLTRAAY 282

Query: 241 HLLGLGTYFTAGEKEVRAWTFKRGIKAPQAASIIHSDFERGFIRAVTMSYDDLIQYGSEK 300
            HLLGLGTYFTAGEKEVRAWTFKRGIKAPQAA IIHSDFERGFIRAVTMSYDDL+ YGSEK
Sbjct: 283 HLLGLGTYFTAGEKEVRAWTFKRGIKAPQAAGIIHSDFERGFIRAVTMSYDDLMTYGSEK 342

Query: 301 AVKEAGRLREEGKEYIVQDGDIMEFRFNV 329
            AVKEAGRLREEGKEY+VQDGDIMEFRFNV
Sbjct: 343 AVKEAGRLREEGKEYVVQDGDIMEFRFNV 371
```

SEQ ID 8876 (GBS177) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 4; MW 41.2kDa).

The GBS177-His fusion product was purified (Figure 118A; see also Figure 202, lane 7) and used to immunise mice (lane 1 product; 20μg/mouse). The resulting antiserum was used for Western blot, FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1657

A DNA sequence (GBSx1752) was identified in *S.agalactiae* <SEQ ID 5133> which encodes the amino acid sequence <SEQ ID 5134>. This protein is predicted to be stage V sporulation protein C (pth). Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2212(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10301> which encodes amino acid sequence <SEQ ID 10302> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB03787 GB:AP001507 stage V sporulation protein C
            (peptidyl-tRNA hydrolase) [Bacillus halodurans]
```

```
 Identities = 89/187 (47%), Positives = 127/187 (67%), Gaps = 2/187 (1%)

Query: 6   VKMIVGLGNPGSKYNDTKHNIGFMAVDRIVKDLDVNFTEDKNFKAEIGSDFINGEKIYFI 65
           +K+IVGLGNPG+KY+ T+HN+GF  VD + + L++   + K     G   I+GEKI+ +
Sbjct: 1   MKLIVGLGNPGAKYDGTRHNVGFDVVDAVARRLNIEIKQSKA-NGLYGEGRIDGEKIFLL 59

Query: 66  KPTTFMNNSGIAVKALLTYYNISIKDMIIIYDDLDMEVGKIRFRQKGSAGGHNGIKSIIA 125
           KP TFMN SG +V+  L YYN+ ++D+++IYDDLD+ VGKIR RQKGSAGGHNG+KS+IA
Sbjct: 60  KPQTFMNRSGGSVRPFLKYYNMEVEDLLVIYDDLDLPVGKIRLRQKGSAGGHNGMKSLIA 119

Query: 126 HLGTQEFDRIKVGIGRPNGRMTVINHVLGKFDKNDEIMLLMTLDKVDNAVNYYLQTNDFQ 185
           HLGT +F RI+VG+ RP     TV+ HVLG++    ++ I  +D    A  + +   F
Sbjct: 120 HLGTSDFKRIRVGVDRPAGETVVQHVLGRYRPEEKDAISEAIDLSAEAARAFTK-KPFL 178

Query: 186 KIMQKYN 192
           + M  +N
Sbjct: 179 EVMNTFN 185
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5135> which encodes the amino acid sequence <SEQ ID 5136>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2840(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 148/189 (78%), Positives = 166/189 (87%)

Query: 5   MVKMIVGLGNPGSKYNDTKHNIGFMAVDRIVKDLDVNFTEDKNFKAEIGSDFINGEKIYF 64
           MVKMIVGLGNPGSKY  TKHNIGFMA+D IVK+LDV FT+DKNFKA+IGS FIN EK+YF
Sbjct: 16  MVKMIVGLGNPGSKYEKTKHNIGFMAIDNIVKNLDVTFTDDKNFKAQIGSTFINEKVYF 75

Query: 65  IKPTTFMNNSGIAVKALLTYYNISIKDMIIIYDDLDMEVGKIRFRQKGSAGGHNGIKSII 124
           +KPTTFMNNSGIAVKALLTYYNI I D+I+IYDDLDMEV K+R R KGSAGGHNGIKSII
Sbjct: 76  VKPTTFMNNSGIAVKALLTYYNIDITDLIVIYDDLDMEVSKLRLRSKGSAGGHNGIKSII 135

Query: 125 AHLGTQEFDRIKVGIGRPNGRMTVINHVLGKFDKNDEIMILNTLDKVDNAVNYYLQTNDF 184
           AH+GTQEF+RIKVGIGRP    MTVINHV+G+F+  D I I  TLD+V NAV +YLQ NDF
Sbjct: 136 AHIGTQEFNRIKVGIGRPLKGMTVINHVMGQFNTEDNIAISLTLDRVVNAVKFYLQENDF 195

Query: 185 QKTMQKYNG 193
           +KTMQK+NG
Sbjct: 196 EKTMQKFNG 204
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1658

A DNA sequence (GBSx1753) was identified in *S.agalactiae* <SEQ ID 5137> which encodes the amino acid sequence <SEQ ID 5138>. This protein is predicted to be transcription-repair coupling factor (mfd). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2456(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD03810 GB:AF054624 transcription-repair coupling factor
          [Lactobacillus sakei]
 Identities = 523/1051 (49%), Positives = 733/1051 (68%), Gaps = 20/1051 (1%)

Query: 1    MNIIELFSQNKVVRTWHSGLVTNSRQLVMGPSGASKAIASAYEKLSKKIMVVTATQTD 60
            M++I +    + V++          RQL+ G SG++K + +A+ Y++   + ++++ +
Sbjct: 1    MDLISMLGNTQQVQSVLENQKPGVRQLLTGLSGSAKTLFLATIYKQQRQPLLIIESNMPQ 60

Query: 61   SDKLSSDISSLIGEDNVYQFPADDVPAAERFIPSSLDKSISRLSALRFLKDPEKNGVLITS 120
            +++++ D+++ +  D +Y F  ++V AAE   SS +      R+  L FL   +K G+++TS
Sbjct: 61   ANQVAEDLANQLNGDQIYTFPVKEVMAAEIAVSSPESRAKRVRTLSFLATGKK-GIVVTS 119

Query: 121  ISGLRLLLPNPEVFSKSQYKFEIGQECVLDKLCKNLVNLGVQKVSQVFSPGEFSQRGDIL 180
            ++G+R LLP     + SQ + E+G E     L   L  +GY +    V  PGEF+ RGDI+
Sbjct: 120  VAGMRRLLPTVRQWRDSQTQIEMGGEVDFKILGAQLAEMGYHRDKLVGKPGSEFAMRGDII 179

Query: 181  DIFEMTQEYPYRLEFFGDEIDGIRQFDIDTQKSLKQLSSVQISPADDIILQDADFERAKK 240
            DIF +  E P R+E F  E+D IR F+ DTQ+S++  LESV I PA D++     A  E A +
Sbjct: 180  DIFFLDTENPVRIELFDTEVDAIRSFEADTQRSIENLESVAIMPATDLLANAAQLEMAGE 239

Query: 241  KLEG-YIVTASEVQ-----------RTYLSEVLSTTSNHFKHSDIRRFLSIFYEKEWGI 287
              L+  Y  TA+++              T +S +L+     + ++  F+   Y     +
Sbjct: 240  ALQADYQQTAAKITAKDDQKALAVNPETPISRLLAGS----RLENLALPVDYLYPDHTSL 295

Query: 288  LDYIPEGTPLFVDDPQKIVDRNAKLDLEIASLLTEDLQQGKSHSSLNYPSDPYKQLRQYQ 347
            +DY     + DD+ +I +    L  E A+  T+ L   +   +     D +  ++Q Q
Sbjct: 296  IDYFKNSGLVVFDDYPRIQETQRVLAEEAANWQTDMLGSRRLLPAQKLLVDVHHLMKQDQ 355

Query: 348  -PATFFSNFHKGLGNLKFDKLHHFTQNGMQEFFNQFFLLVDEINRYKKSGATVLLQVDSQ 406
             P  + S F KG+G LK D L +    +Q++FF+Q PLL  E++R++K   TV++ V
Sbjct: 356  HPHLYLSLFQKGMGKLKLDTLGNMETRNVQQFFSQMPLLKTENSRWQKQQQTVVVLVSDA 415

Query: 407  KGLNLLQENLKEYGLDLIISDKNDIVQKESQLIVGHLSNGFYFADEKIVLITERBIYHRR 466
            K +  + +   ++ ++  ++ K  +V  + G+ G L NGF    D K+V++TR+E+++
Sbjct: 416  KRVKKIDQTFHDFEIEATVTTKTXLVAGQIQIVQSSLQNGFELPDLKLVVLTEKKLFNTA 475

Query: 467  VKRKIRRSNISNAERLKDYNELSVGDYVVHNVHGVGKPLGIRTIEIQGIHRDYLTIQYQN 526
             K+K+RR  ++NAERLK Y+EL  GDYVVH  HG+G+++G+ET+E+ G+H+DY+TI Y++
Sbjct: 476  PKKKVRRQTLANAERLKSYSELKPGDYVVHVNHGIGEYVGMETLEVDGVHQDYITILYRD 535

Query: 527  ADRISIPVEQIELLTKYVGADGKSPKINTLNDGRPKKAKQRVAKQVEDIADDLLKLVAER 586
             ++ IPV Q++++ KYVSA+ K PKIN L   ++K K +V+ ++EDIADDL++ LVA+R
Sbjct: 536  NGSLPIPVTQLDMVQKYVSAESKTPKINKLGGAEWQKTKSKVSAKIEDIADDLIELVAQR 595

Query: 587  SQLQGFAFSPDDNMQNDFDNDPAYVRTEDQLRSIKEIKQDMGNRPMDRLLVGDVGFGKT 646
               +G+AF  DD +Q DF+N FAY ET+DQLRS  EIK DMR  EPMDRLLVGDVGFGKT
Sbjct: 596  EAEKGYAFPKDDQLQADFENQFAYPETDDQLRSTAEIKHDMEKVEPMDRLLVGDVGFGKT 655

Query: 647  EVAMRAAFKAVNDHKQVVVLVPTTVLAQQHFENFKERPSNYPVTVDVLSRFKSKKEQTDT 706
            EVA+RAAFKAV   KQV  LVPTT+LAQQH+EN   RP+++PV  + +LSRF+++KE T T
Sbjct: 656  EVALRAAFKAVAGKQVAFLVPTTILAQQHYENMLARFADFPVVELGLLSRFKTRKEVTAT 715

Query: 707  LKRLSKGQVDIIIGTHRLLSQDVVFSDLGLIVIDEEQRFGVKHKEKLKKLRTKVDVLTLT 766
            LK L KGQVDI+IGTHRLLS+DVVF DLGL++IDEEQRFGVKHKE+LK+LK +VDVLTLT
Sbjct: 716  LKGLEKGQVDIVIGTHRLLSKDVVFKDLGLLVIDEEQRFGVKHKSKLKQLKQAVDVLTLT 775

Query: 767  ATPIPRTLHMSMLGIRDLSVIETPPTNRYPVQTYVLETNPGLVRSAIIRELDRGGQVFYV 826
            ATPIPRTLHMSMLG+RDLSVIETPPTNRYP+QTYV+R N G +REAI RE++R GQVFY+
Sbjct: 776  ATPIPRTLHMSMLGVRDLSVIETPPTNRYPIQTYVMEQNAGAMREAIERELERNGQVFYL 835

Query: 827  YNKVDTIDQKVSELQELVPEASIGFVHGQMSEIQLENTLIDFINGDYDVLVATTIIETGV 886
            +N+V  I+Q V E+Q LVPEA++G+ HGQM+E QLE  + DF+ G YDVLV TTIIETGV
Sbjct: 836  HNRVSDIEQTVDEIQALVPEATVGYAHGQMTEAQLEGVIYDFVQGKYDVLVTTTIIETGV 895
```

```
Query: 887  DISNVNFTLFVENADHMGLSTLYQLRGRVGRGNRIAYAYALMYRPDKVLTEISEKRLDAIKG 946
            D+ NVNF+ VE+ADH GLS LYQLRGR+GRS+R+AY Y MY+PDKVLTE+SEKRL AIK
Sbjct: 896  DMPNVNFMIVEDADHYGLSQLYQLRGRIGRSSRVAYGYFMYKPDKVLTEVSEKRLQAIKD 955

Query: 947  FTELGSGFKIAMRDLSIRGAGNILGAEQSGFIDSVGFEMYSQLLEQAIATKQGKSLIRQK 1006
            FTELGSGFKIAMRDLSIRGAGN+LG  Q GFIDSVGF++YSQ+L +A+A KQGK  +  K
Sbjct: 956  FTELGSGFKIAMRDLSIRGAGNLLGKQQHGFIDSVGFDLYSQMLSEAVAKKQGKK-VAAK 1014

Query: 1007 GNAELALQIDAYLPARVISDERQKIEIYKRI 1037
            GNAE+ L+++AYLP +YI+D+RQKIEIYKRI
Sbjct: 1015 TNAEILLKLEAYLPDDYINDRQRQIEIYKRI 1045
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5139> which encodes the amino acid sequence <SEQ ID 5140>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2826(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 875/1161 (75%), Positives = 1032/1161 (88%)

Query: 1    MNIIELFSQNKVVRTWHSGLVTNSRQLVWGPSGASKAIAISAYEKLSKKIMVVTATQTD 60
            M+I+ELFSQNK V++WHSGL T  RQLVWG SG+SK +AIASAY    KKI+VVT+TQ +
Sbjct: 1    MDILELFSQNKKVQSWHSGLTTLGRQLVWGLSGSSKTLAIASAYLDDQKKIVVVTSTQNE 60

Query: 61   SDKLSSDISSLIGEDNVVQFFADDVPAAEFIFSSLDKSISRLSALRFLKDPEKNGVLITS 120
             +KL+SD+SSL+ E+ V+QFFADDV AAEFIF+S+DK+ +SR+  L+PL++P+  GVLI S
Sbjct: 61   VEKLASDLSSLLDEELVFQFFADDVAAAEFIFASMDKALSKIETLQPLRNPKSQGVLIVS 120

Query: 121  ISGLRLLLPNPEVFSKSQYKFEIGQECVLDKLCKNLVNLGVQKVSQVFSPGEFSQRGDIL 180
            +SGLR+LLPNP+VF+KSQ +  +G++   D L K L+  +GVQKVSQV SPGEFS+RGDIL
Sbjct: 121  LSGLRILLPNPDVFTKSQIQLTVGEDYDSDTLTKQLMTIGVQKVSQVISPGEFSRRGDIL 180

Query: 181  DIFEMTQEYPYRLEFFGDEIDGIRQFDIPTQKSLKQLESVQISPADDIILQDADFERAKK 240
            DI+E+TQE PYRLEFFGD+ID IRQF  +TQKS +QLE + I+PA D+I  +DF+R  +
Sbjct: 181  DIYEITQELPYRLEFFGDDIDSIRQFHPETQKSFEQLEGIFINPASDLIFEVSDFQRGIE 240

Query: 241  KLEGYLVTASEVQKTYLSEVLSTTSNHFKHSDIRRFLSIFYKEWGILDYIPEGTPLFVD 300
            +LE  L TA + +++YL +VL+ ++N FKH DIR+F S+FYKEW +LDYIP+GTP+F D
Sbjct: 241  QLEKALQTAQDDKKSYLEDVLAVSKNGFKHDIRKPQSLFYKEWSLLDYIPNGTPIFFD 300

Query: 301  DPQKIVDRNAKLDLEIASLLTEDLQQGKSHSSLNYFSDPYKQLRQYQPATPFSNPHKGLG 360
            DPQK+VD+NA+ DLEIA+LLTEDLQQGK+ S+LNYF+D Y++LR Y+PATPFSNPHKGLG
Sbjct: 301  DPQKLVDKNARFDLEIANLLTEDLQQGKALSNLNYFTDNYRELRHYKPATPFSNPHKGLG 360

Query: 361  NLKFDKLHHPTQYQMQEFFNQPDLLVDRINRYKKSGATVLLQVDSQKGLNLLQENLKEYG 420
            N+KFD++H  TQY MQEFFNQPDLL+DEI EY+K+  TV++QV+SQ      L+++ ++Y
Sbjct: 361  NIKFDQMHQLTQYAMQEFFNQPDLLIDEIKRYQKNQTFVIVQVESQYAYERLEKSFQDYQ 420

Query: 421  LDLIISDXNDIVQKESQLIVGHLSNGFVFADEKIVLITRETYHRRVKRKIRRSNISNAE 480
              L +   N IV +ESQ+++G +S+GFYFADEK+ LITE EIYH+++KR+ RRSNISNAE
Sbjct: 421  FRLPLVSANQIVSKRESQIVIGAISSGFYFADEKLALITHEIYHKKIKRRARRSNISNAE 480

Query: 481  RLKDYNELSVGDYVVHNVHGVGKPLGIRTIRIQGIHRDYLTIQYQNADRISIPVEQIELL 540
            RLKDYNELVGDYVVHNVHG+G+PLGIRTI+IQGIHRDY+TIQYQN+DRIS+P++QI  L
Sbjct: 481  RLKDYNELAVGDYVVHNVHGIGRPLGIRTIQIQGIHRDYVTIQYQNSDRISLPIDQISEL 540

Query: 541  TKYVSADGKEPKINTLNDGRFKKAKQRVAKQVEDIADDLLKLYAERSQLQSFAFSPDDNM 600
            +KYVSADGKEPKIN LNDGRF+K KQ+VA+QVEDIADDLLKLYAERSQ +GF+FSPDD++
Sbjct: 541  SKYVSADGKEPKINKLNDGRFQKTKQKVARQVEDIADDLLKLYAERSQQKGFSFSPDDDL 600
```

```
Query: 601  QNDPDNDFAYVETEDQLRSIKEIKQDMEGNRPMDRLLVGDVGFGKTEVAMRAAFKAVNDH 660
            Q  PD+DFA+VETEDQLRSIKEIK DME  +PMDRLLVGDVGFGKTEVAMRAAFKAVNDH
Sbjct: 601  QRAPDDDFAFVETEDQLRSIKEIKADMESMQPMDRLLVGDVGFGKTEVAMRAAFKAVNDH 660

Query: 661  KQVVVLVPTTVLAQQHPENFKSRPSNYPVTVDVLSRFRSKKBQTDTLKRLSKGQVDIIIG 720
            KQV VLVPTTVLAQQH+ENFK RP NYPV VDVLSRFRSKKBQ +TL+R+ KGQ+DIIIG
Sbjct: 661  KQVAVLVPTTVLAQQEYENFKARPENYPVEVDVLSRFRSKKBQAETLERVRKGQIDIIIG 720

Query: 721  THRLLSQDVVFSDLGLIVIDEBQRFGVKHKEKLKELKTKVDVLTLTATPIPRTLHMSMLG 780
            THRLLS+DVVFSDLGLIVIDEBQRFGVKHKR LKELKTKVDVLTLTATPIPRTLHMSMLG
Sbjct: 721  THRLLSKDVVFSDLGLIVIDEBQRFGVKHKETLKELKTKVDVLTLTATPIPRTLHMSMLG 780

Query: 781  IRDLSVIETPPINRYPVQTYVLETNGLVRKAIIREIDRGGQVFFVYNKVDTIDQKVSEL 840
            IRDLSVIETPPINRYPVQTYVLE NPGLVRKAIIRE+DRGGQ+FFVYNKVDTI++KV+EL
Sbjct: 781  IRDLSVIETPPINRYPVQTYVLENNPGLVRKAIIREMDRGGQIFFVYNKVDTIEKKVAEL 840

Query: 841  QELVPEASIGFVHGQMSEIQLENTLIDFINGDYDVLVATTIIETGVDISNVNTLFVENAD 900
            QELVPEASIGFVHGQMSEIQLSNTLIDFINGDYDVLVATTIIETGVDISNVNTLFIENAD
Sbjct: 841  QELVPEASIGFVHGQMSEIQLSNTLIDFINGDYDVLVATTIIETGVDISNVNTLFIENAD 900

Query: 901  HMGLSTLYQLRGRVGRSNRIAYAYLMYRPDKVLTEISEKRLDAIKGFTELGSGFKIAMRD 960
            HMGLSTLYQLRGRVGRSNRIAYAYLMYRPDKVLTE+SEKRL+AIKGFTELGSGFKIAMRD
Sbjct: 901  HMGLSTLYQLRGRVGRSNRIAYAYLMYRPDKVLTEVSEKRLEAIKGFTELGSGFKIAMRD 960

Query: 961  LSIRGAGNILGASQSGFIDSVGFEMYSQLLEQAIATKQSKSLIRQKGNAELALQIDAYLP 1020
            LSIRGAGNILGASQSGFIDSVGFEMYSQLLEQAIA+KQSK+ +RQKGN E+ LQIDAYLP
Sbjct: 961  LSIRGAGNILGASQSGFIDSVGFEMYSQLLEQAIASKQSKTTVRQKGNTEINLQIDAYLP 1020

Query: 1021 AEYISDERQKIEIYKRIRELETRADYEALQDELIDRFGEYPDQVAYLLEIGLLKAYLDLA 1080
               +YI+DERQKI+IYKRIRE+++R DY  LQDELD+DRFGEYPDQVAYLLEI LLK Y+D A
Sbjct: 1021 DDYIADERQKIDIYKRIRELQSREDYLNLQDELMDRFGEYPDQVAYLLEIALLKHYMDNA 1080

Query: 1081 FTELVERKGNEISILFEKASLKYFLTQDYFEALSKTQLKARISETNKHMEVVFNIKHKN 1140
               F ELVERK N++ + FE  SL YFLTQDYFEALSKT LKA+ISE  GK+++VF+++H+K+
Sbjct: 1081 FAELVERKNNQVIVRFEVTSLTYFLTQDYFEALSKTHLKAKISEHQGKIDIVFDVRHQKD 1140

Query: 1141 YEIIEELLKPAECFIEIKSRK 1161
               Y I+EEL+ F E   EIK RK
Sbjct: 1141 YRILEELMLPGERLSEIKIRK 1161
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1659

A DNA sequence (GBSx1754) was identified in *S.agalactiae* <SEQ ID 5141> which encodes the amino acid sequence <SEQ ID 5142>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4347(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11835 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 50/84 (59%), Positives = 70/84 (82%)

Query: 1  MRLDKYLKVSRIIKRRPVAKEVADKGRVKVNGVLAKSSTDLKLNDQVEIRFGNKLLTVKV 60
          MRLDK+LKVSR+IKRR +AKEVAD+GR+ +NG  AK+S+D+K  D++ +RFG KL+TV+V
Sbjct: 1  MRLDKFLKVSRLIKRRTLAKEVADQGRISINGNQAKASSDVKPGDELTVRGQKLVTVQV 60

Query: 61 LEMKDSTKKEDAIKMYEIINETRI 84
```

```
          E+KD+TKKE+A  MY I+ E ++
Sbjct: 61 NELKDTTKKEEAANMYTILKEEKL 84
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5143> which encodes the amino acid sequence <SEQ ID 5144>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2963(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 72/90 (80%), Positives = 85/90 (94%)

Query: 1  MRLDKYLKVSRIIKRRPVAKEVADKGRVKVNGVLAKSSTDLKLNDQVEIRFGNKLLTVKV 60
          MRLDKYLKVSR+IKRR VAKEVADKGR+KVNG+LAKSST++KLND +EI FGNKLLTV+V
Sbjct: 9  MRLDKYLKVSRLIKRRSVAKEVADKGRIKVNGILAKSSTNIKLNDHIEISPGNKLLTVRV 68

Query: 61 LEMKDSTKKEDAIKMYEIINETRIETDEQA 90
          +E+KDSTKKEDA+KMYEII+ETRI  +E+A
Sbjct: 69 IEIKDSTKKEDALKMYEIISETRITLNEEA 98
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1660

A DNA sequence (GBSx1755) was identified in *S.agalactiae* <SEQ ID 5145> which encodes the amino acid sequence <SEQ ID 5146>. This protein is predicted to be DivIC homolog. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.12   Transmembrane   34 -  50 (  31 -  56)

----- Final Results -----
                bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC98903 GB:AF023181 DivIC homolog [Listeria monocytogenes]
 Identities = 36/119 (30%), Positives = 65/119 (54%), Gaps = 2/119 (1%)

Query: 2   SKFNVVQLNNQVINDE-NLKKRYEAESLRRKNRLMGWVLIFVMLLFLLPTYNLVKSYRTL 60
           +K  V ++ N+YI D   +KK      +    RL     +IF ++  +L T    K   TL
Sbjct: 4   AKSKVARIENRYIKDTATMKKTRSRRRIALERRLAFMAIIFAVVGGLL-TITYTKQVLTL 62

Query: 61  QERRQEVVKLTKDYQTLTNRTENQKLLAKQLKNPDYVQKYARAKYYPSKTGEMIYPLPD 119
           +E++++ V++ K     + +  ++      K+L N DY+ K AR++YY SK GE+I+ +P+
Sbjct: 63  KEKKEEQVQVDKKMVAMKDEQDSLNEQIKKLINDDYIAKLARSEYYLSKDGEIIFNIPE 121
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5147> which encodes the amino acid sequence <SEQ ID 5148>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence
```

```
INTEGRAL    Likelihood = -3.93   Transmembrane   34 -  50 (  32 -  51)

----- Final Results -----
                bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC98903 GB:AF023181 DivIC homolog [Listeria monocytogenes]
 Identities = 27/116 (23%), Positives = 59/116 (50%)

Query: 3   KPSIVQLNNHYIKKKNLKKKFKHEESQKRNRMGWILVSMMFLFILPTYNLVKSYVDFEK 62
            K  + ++ N YIK     KK          R + ++ +     +  L T     K  +  ++
Sbjct: 5   KSKVARIRNRYIKDTATMKKTRSRRRIALFRRLAFMAIIFAVVGGLLTITYTKQVLTLKS 64

Query: 63  QNQQVVKLKKEYNELSESTKKEKQLARRLKDDNFVKKYARAKYYLSREGEMIYPIP 118
            + ++ V++ K+   + +       +  ++L +D+++ K AR++YYLS++GE+I+ IP
Sbjct: 65  KKEKQVQVDKKMVAMKDEQDSLNEQIKKLHNDDYIAKLARSEYYLSKDGEIIFNIP 120
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 73/123 (59%), Positives = 99/123 (80%)

Query: 1   MSKPNVVQLNNQYINDENLKKRYEAEELRRKNRLMGWVLIFVMLLFILPTYNLVKSYRTL 60
            M KP++VQLNN YI  ENLKK++E EE +++NR MGW+L+ +M LFILPTYNLVKSY
Sbjct: 1   MKKPSIVQLNNHYIKKENLKKKFEEEESQKRNRFMGWILVSMMFLFILPTYNLVKSYVDF 60

Query: 61  QERRQSVVKLTKDYQTLTNRTENQKLLAKQLWNPDVVQKYARAKYYFSKTGEMIYPLPDL 120
            +++ Q+VVKL K+Y  L+  T+ +K LA++LK+ ++V+KYARAKYY S+ GEMIYP+P L
Sbjct: 61  EKQNQQVVKLKKEYNELSESTKKEKQLARRLKDDNFVKKYARAKYYLSREGEMIYPIPGL 120

Query: 121 LPK 123
             LPK
Sbjct: 121 LPK 123
```

SEQ ID 5146 (GBS418) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 3; MW 42kDa).

GBS418-GST was purified as shown in Figure 219, lane 4-5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1661

A DNA sequence (GBSx1756) was identified in *S.agalactiae* <SEQ ID 5149> which encodes the amino acid sequence <SEQ ID 5150>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4355(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1662

A DNA sequence (GBSx1757) was identified in *S.agalactiae* <SEQ ID 5151> which encodes the amino acid sequence <SEQ ID 5152>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.52   Transmembrane    4 -  20 (   3 -  22)

----- Final Results -----
                bacterial membrane --- Certainty=0.3208(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5153> which encodes the amino acid sequence <SEQ ID 5154>. Analysis of this protein sequence reveals the following:

```
      Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 205/428 (47%), Positives = 285/428 (65%)

Query: 1   MKKVLTFLLCSLYFVSIPAISTEEPLTLSQNRRYALTQTVDKEMYFDAIPERPTTKIEI 60
           M+K+L  +L + +    +P ISTE+ L  S+N  Y L Q VV     +++ IP  P    E
Sbjct: 1   MRKLLAAMLMTTFFLTPLPVISTEKKLIFSKNAVYQLKQDVVQSTQFYNQIPSNPNLYQET 60

Query: 61  SSFQDEALTITGETLVFNTLLSIVSLTINSNGIPVFTLSNGQPIKASREAIFNDLVSKQQ 120
            +++D  LT+    L  N  L I SL +N    +PVF  L++G +++A+R+  I++D+V  Q
Sbjct: 61  CAYKDSDLTLPAGRLGVNQPLLIKSLVLNKESLPVFSLADGTYVEANRQLIYDDIVLNQV 120

Query: 121 SVSLDYWLKPSFVTVEAPYTNGVSEVGNNLKPYSRVHLVEQARTEHGIYYKTDSGFWISV 180
            +   +W +     Y APY  G   + ++     +VH  + A+T HG YY  D   W S
Sbjct: 121 DIDSYFWTQKKLRLYSAPYVLGTQTIPSSFLFAQKVHATQMAQTNHGTYYLIDDKGWASQ 180

Query: 181 EDLSVADNRMAKVQEVLLEKYNKDKYGIYIKQLNTQTVAGINIDRSMYSASIAKLATLYA 240
            EDL   DNRM KVQE+LL+KYN   Y I++KQLNTQT AGIN D+ MY+ASI+KLA LY
Sbjct: 181 EDLVQFDNRMLKVQRMLLQKYNNPNYSIPVKQLNTQTSAGINADKMYAASISKLALPLYI 240

Query: 241 SQEQVKLGKLSLDSKFKYKDNVNQFPNSYDPSGSGKLEKKADHKLYTVKELLEATAKESD 300
             Q+Q++  XL+ +    Y  +VN F   YDP GSGK+ K AD+K Y V++LL+A A++SD
Sbjct: 241 VQKQLQKKXLAENKTLTYTKDVNHFYGDYDPLGSGKISKIADNKDYRVEDLLKAVAQQSD 300

Query: 301 NVATNMLGYYVNQYDSMPQTQVDTISGMHWDMKKRQISPQAAGKMMEAIYYQNGDIVNY 360
            NVATN+LGYY+ +QYD  F++++  +SG+ WDM++R ++ ++A  MMEAIY+Q G I++Y
Sbjct: 301 NVATNILGYYLCHQYDKAPRSEIKALSGIDWDMEQRLLTSRSAANMMEAIYHQKGQIISY 360

Query: 361 LSKTDFDNTRIPKNIPVKVAHKIGDAYDYKHDAAIVYABQPFIMIIPTDKSSYDDITKIA 420
            LS T+FD  RI KNI V VAHKIGDAYDYKHD AIVY   PFI+ IPT+KS+Y+DIT IA
Sbjct: 361 LSNTKFDQQRITKNITVPVAHKIGDAYDYKHDVAIVYGNTPFILSIPTNKSTYRDITAIA 420

Query: 421 DDVYQVLK 428
            DDVY +LK
Sbjct: 421 DDVYGILK 428
```

SEQ ID 5152 (GBS116) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 3; MW 48.5kDa). The GBS116-His fusion product was purified (Figure 202, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 316), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1663

A DNA sequence (GBSx1758) was identified in *S.agalactiae* <SEQ ID 5155> which encodes the amino acid sequence <SEQ ID 5156>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2260(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD35664 GB:AE001733 conserved hypothetical protein [Thermotoga maritima]
 Identities = 100/404 (24%), Positives = 181/404 (44%), Gaps = 61/404 (15%)

Query: 19  QKVLIAVSGGIDSINLLQFLYQVQKELSISIGIAHINHGQRKESEKEEEYIRQWQQIHDV 78
           + VL+AVSGGIDS+ LL  L ++   L I I  AH++H  R+ S ++ E++ +  +  ++
Sbjct: 6   EHVLVAVSGGIDSMTLLYVLRKFSPLLKIKITAAHLDHRIESSSRRDREPVERICRQWNI 65

Query: 79  PVFISYF--------QGIFSKDRARNHRVNPFSKVMREGGYTALVTAHHADDQAETVFMR 130
            PV  S          G   E+ AR  RY+F  +  ++ G + +  AHH +D  ETV  R
Sbjct: 66  PVETSEVDVPSLNKDSGKTLEEIAREVRYDPLKRTAKKVGASKIALAHHKNDLLETVVHR 125

Query: 131 ILRGSRLRYLSGIKQVSAPANGQLIRPFLEYKKELLP------NIPHPEDASNASSDYLR 184
           ++RG+    L+ I          + IRPFL +K+  +       N+ +  D +N +  Y R
Sbjct: 126 LIRGTGPLGLACISP----KREEFIRPFLVFKRSEIEEYARKNNVPYVVDETNYNVKYTR 181

Query: 185 NRIRNVYFPALERENNQLKDSLITLSEETCLPTALTDLTRSIEVTNCYDF--------- 235
            N IR+   P ++  N  ++D++  L   T  L   +     N Y +
Sbjct: 182 NFIRHRIVPLMKELNPTVEDAVYRLVSVTHLLRNFVERTVQDFVERNVYFYKDYAVFVEP 241

Query: 236 --LRQTHSVQEFLLQDYISKFPDLQVHKBQPRVILKLIRTKANIDYTIKSGYFLHKDYES 293
             L     V  ++L++   + P+ +          KLI T  +  + SG F+ + +
Sbjct: 242 EDLPLFLEVTRWVLKRMYGRVPEYE----------KLIGTLKSKRVELWSGIFVERSFGY 291

Query: 294 FEITKIHPKTDSPKVEKRLKLHNIQIPSQYLFSYGKFISQADITIPIYDT---SPIILRR 350
              + K       FK + R+E+             G  +      I + +        +R
Sbjct: 292 VAVGK-----TVFKKKYRVEVK-----------GDMLEMEGFKIRVVNNRNDMKFWVRN 334

Query: 351 RKKGDRIFLGNHTKKIRKLFIDEKIT--LKERKKAVIGKQNKEL 392
            RKKGDRI +    +K++ +FI++K+    ++R   ++ E+++ L
Sbjct: 335 RKEGDRIIVNGRRKKLKDVFIEKKVPTFYRDRVPLLVDEHDRVL 378
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5157> which encodes the amino acid sequence <SEQ ID 5158>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2187(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 218/424 (51%), Positive = 290/424 (67%), Gaps = 2/424 (0%)

Query: 2   YNTILKDTLSKGLFTAHQKVLIAVSGGIDSINLLQFLYQYQKELSISIGIAHINHGQRKE 61
           Y  I  +  +K  F  H+ VLIAVSGG+DS+NLL FLY +Q +L I IGIAH+NH QR E
Sbjct: 4   YQEIFNEIKNKAYFKNHRHVLIAVSGGVDSMNLLHFLYLFQDKLKRIGIAHVNHQQRSS 63

Query: 62  SEKEEEVIRQNGQIHDVPVFISYPQQIFSEDRARNHRYNFFSKVMREEGYTALVTAHHAD 121
           S+ EE Y++ W +  HD+P++++S F+GIFSE  AR+ EY FF  +M +  Y+ALVTAHH+D
Sbjct: 64  SDSEEAYLKCWAKKHDIPIYVSNFEGIFSESGKARDWRYAFFKSIMLKNNYSALVTAHHSD 123

Query: 122 DQAETVFMRILRGSRLRYLSGIKQVSAFANCQLIRPFLPYKKELLDWIPHFEDASNASSD 181
           DQAET+ MR++RGSRLR+LSGIK V  FANGQLIRPFL + K+ LP IPHFED+SN
Sbjct: 124 DQAETILMKLIRGSRLRHLSGIKSVQPFANGQLIRPFLTFSKKDLPEIPHFEDSSNRELS 183

Query: 182 YLRNRIRNVYFPALERENNQLKDSLITLSEETECLFPALTLDLTRSIEVTNCYDFLRQTHS 241
           +LRNR+RN Y P  L++EN +    L  L+ E   LF A  +LT  I  T+  +F  Q+ S
Sbjct: 184 FLRNRVRNNYLPLLKQENPRFIQGLNQLALENSLLFQAPKELTNHITTTDLTEFNEQSKS 243

Query: 242 VQEFLLQDYISKFPDLQVSKEQFRVILKLIRTKANIDYTIKSGYFLHKDYESFHITKIHP 301
           +Q FLLQDY+  FPDL + K QF  +L++I+T     Y +K  Y++  D  SF ITKI P
Sbjct: 244 IQYFLLQDYLSGFPDLDLKKSQFTQLLQIIQTAKQGYYYLKKDYYIFIDKFSFKITKIVP 303

Query: 302 KTDSFKVEKRLELHNIQIFSQYLFSY--GKFISQADITTPIYYDTSPILLRRRKEGDRIFL 359
           KT+  K EK LE  +  +  Y FS+        Q  ++IP++  S I LR R+ GD I
Sbjct: 304 KTELVKEEKMLEVDSKLCYRDYYFSPMPKENEDQGQVSIPLFSLSSIKLRSRQSGDYISF 363

Query: 360 GNFTKKIRRLPIDEKITLKERESAVIGEQNKELIFVIVAGRTYLRKDSSHDIMKGKLYIK 419
           G+ ++KIRRLPIDEK T+ ER+ A+IGEQ++++IFV++  +TYLRK  +HDIM  KLYI+
Sbjct: 364 GHFSKKIRRLPIDEKFTIAERQNAIIGEQDEQIIFVLIGNKTYLRKACKHDIMLAKLYID 423

Query: 420 NLEK 423
             LEK
Sbjct: 424 KLEK 427
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1664

A DNA sequence (GBSx1759) was identified in *S.agalactiae* <SEQ ID 5159> which encodes the amino acid sequence <SEQ ID 5160>. This protein is predicted to be hypoxanthine-guanine phosphoribosyltransferase (hpt). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane   37 -  53 (  37 -  53)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA48876 GB:X69123 hypoxanthine guanine
          phosphoribosyltransferase [Lactococcus lactis]
 Identities = 121/179 (67%), Positives = 152/179 (84%), Gaps = 1/179 (0%)

Query: 2   LENDIKKVLYSEEDIILKTKELGAKLTADYAGKNPLLVGVLKGSVPFMAELLKHIDTHVE 61
           L+  I+KVL SEE+II K+KELG  LT +Y GKNPL++G+L+GSVPF+ABL+KHID H+E
Sbjct: 6   LDKAIEKVLVSEEEIIEKSKELGEILTKEYEGKNPLVLGILRGSVPFLAELIKHIDCHLE 65
```

```
Query: 62  IDFMVVSSYHGGTTSSGEVKLLKDVDTNIEGRDVIFIEDIIDTGRTLKYLRDMFKYRQAN 121
           DFM VSSYHGGT SSGEVK++ DVDT ++GRD++ +EDIIDTGRTLKYL+++ ++R AN
Sbjct: 66  TDFMTVSSYHGGTKSSGEVKLILDVDTAVKGRDILIVEDIIDTGRTLKYLKELLEHRGAN 125

Query: 122 SVKVATLFDKPEGRLVDIDADYVCYDIPNEFIVGFGLDYAENYRNLPYVGVLKEEIYSK 180
           VK+ TL DKPEGR+V+I  DY  + IPNEF+VGFGLDY ENYRNLPYVGVLK E+Y+K
Sbjct: 126 -VKIYTLLDKPEGRIVEIKPDYSGFTIPNEFVVGFGLDYKENYRNLPYVGVLKPEVYNK 183
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5161> which encodes the amino acid sequence <SEQ ID 5162>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4095(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 153/180 (85%), Positives = 171/180 (95%)

Query: 1   MLENDIKKVLYSEEDIILKTKELGAKLTADYAGKNPLLVGVLKGSVPFMAELLKHIDTHV 60
           MLE DI+K+LYSE DII KTK+LG +LT DY  KNPLL+GVLKGSVPFMAKLLKHIDTHV
Sbjct: 1   MLEQDIQKILYSENDIIRKTKKLGEQLTKDYQKKNPLMIGVLKGSVPFMAELMKHIDTHV 60

Query: 61  EIDFMVVSSYHGGTTSSGEVKILKDVDTNIEGRDVIFIEDIIDTGRTLKYLRDMFKYRQA 120
           EIDFMVVSSYHGGT+SSGEVKILKDVDTNIEGRD+I +EDIIDTGRTLKYLRDMFKYR+A
Sbjct: 61  EIDFMVVSSYHGGTSSSGEVKILKDVDTNIEGRDIIIVEDIIDTGRTLKYLRDMFKYRKA 120

Query: 121 NSVKVATLFDKPEGRLVDIDADYVCYDIPNEFIVGFGLDYAENYRNLPYVGVLKEEIYSK 180
           N++K+ATLFDKPEGR+V I+ADYVCY+IPNEFIVGFGLDYAENYRNLPYVGVLKEE+YSK
Sbjct: 121 NTIKIATLFDKPEGRVVKIEADYVCYNIPNEFIVGFGLDYAENYRNLPYVGVLKEEVYSK 180
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1665

A DNA sequence (GBSx1760) was identified in *S.agalactiae* <SEQ ID 5163> which encodes the amino acid sequence <SEQ ID 5164>. This protein is predicted to be cell division protein FtsH (ftsH). Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.11   Transmembrane  139 - 155 ( 133 - 158)
   INTEGRAL    Likelihood = -4.62   Transmembrane    8 -  24 (   7 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.3845(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC16243 GB:AF061748 cell division protein FtsH [Streptococcus pneumoniae] (ver 2)
 Identities = 490/652 (75%), Positives = 561/652 (85%), Gaps = 5/652 (0%)

Query: 5   KNNGFLKNSFIYILLIIAVITTFQYYLKGTSSQ-NQQISYTKLVKQLKAGEIKSISYQPS 63
           +NNG +KN F+++L I  ++T FQY+  G +S  +QQI+TY+LV+++  G +K ++YQP+
Sbjct: 4   QNNGLIKNPFLWLLFIFPLVTGFQYFYSGNNSGGSQQINYTELVQETTDGNVKELTYQPN 63
```

```
Query: 64  GGVVEVSGTYKKATIKSANSFTFLGGSVATKVTGFNSVILPNDSSIKSLVSAAEENNTN 123
           G V+EVSG YK  KT K      F   SV TKV  F S  ILP D+++  L   A ++
Sbjct: 64  GSVIEVSGVYKNPKTSKEGTGIQFPTPSV-TKVEKFTSTILPADTTVSELQKLATDHKAE 122

Query: 124 IQVKHESSSGTWISYIASFLPLVIMIGFFMMMMNQGGGGGARGAMSFGKNKARSSSKDEV 183
           + VKHESSSG WI+ + S +P  I+  F   MM    GGG  R  MSFG++KA+++K+++
Sbjct: 123 VTVKHESSSGIWINLLVSIVPFGILFFFLPSMMGKMGGGNGRNPMSFGRSKAKAANKEDI 182

Query: 184 KVRFSDVAGAKEEKQELIEVVDFLKDPKRYKSLGARIPAGVLLSGPPGTGKTLLAKAVAG 243
           KVRFSDVAGAKEEKQELVEVVDFLKDPKR+  LGARIPAGVLLSGPPGTGKTLLAKAVAG
Sbjct: 183 KVRFSDVAGAKEEKQELVEVVDFLKDPKRFTKLGARIPAGVLLSGPPGTGKTLLAKAVAG 242

Query: 244 EAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAERAIIFIDEIDAVGRRRGAGMGGG 303
           EAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKA  AIIFIDEIDAVGR+RG G+GGG
Sbjct: 243 EAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAAPAIIFIDEIDAVGRQRGVGLGGG 302

Query: 304 NDEREQTLNQLLIEMDGFEGNESIIVIAATNRSDVLDPALLRPGRFDRKVLVGQPDVKGR 363
           NDEREQTLNQLLIEMDGFEGNE IIVIAATNRSDVLDPALLRPGRFDRKVLVG+PDVKGR
Sbjct: 303 NDEREQTLNQLLIEMDGFEGNEGIIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDVKGR 362

Query: 364 EAILRVEAKNKPLADNVDLKVVAQQTPGFVGADLENVLNEAALVAARRNKKVIDASDIDE 423
           EAIL+VEAKNKPLA++VDLK+VAQQ TPGFVGADLENVLNEAALVAARRNK +IDASDIDE
Sbjct: 363 EAILKVEAKNKPLAEDVDLKLVAQQTPGFVGADLENVLNEAALVAARRNKSIDASDIDE 422

Query: 424 AEDRVLAGPSKKDRTISERERRAMVAYHEAGHTIVGLILSNARVVHKVTIVPRGRAGGYMI 483
           AEDRVIAGPSKKD+T+S++ER +VAYHEAGHTIVGL+LSNARVVEKVTIVPRGRAGGYMI
Sbjct: 423 AEDRVIAGPSKKDKTVSQKERELVAYHEAGHTIVGLVLSNARVVEKVTIVPRGRAGGYMI 482

Query: 484 ALPKEDQMLLSKDDMKEQLAGLMGGRVAEEIIFNAQTTGASNDFEQATAMARAMVTEYGM 543
           ALPKEDQMLLSK+DMKEQLAGLMGGRVAEEIIFN QTTGASNDFEQAT MARAMVTEYGM
Sbjct: 483 ALPKEDQMLLSKEDMKEQLAGLMGGRVAEEIIFNVQTTGASNDFEQATQMARAMVTEYGN 542

Query: 544 SEKLGFVQYEGNHAMMAGQMSPEKSYSAQTAQLIDDEVRHLLNEARNKAADIINENRDTH 603
           SEKLGFVQYEGNHAM+ G  SP+KS S QTA   ID+EVR LLNEARNKAA+II  NR+TH
Sbjct: 543 SEKLGFVQYEGNHAMI-GAQSPQKSISEQTAYEIDEEVRSLLNEARNKAAEIIQSNRSTH 601

Query: 604 KLIAEALLKYETLDAAQIKSIFETGKMPETENDEDKARALSYDEIKEKMQEE 655
           KLIAEALLKYETLD+ QIK+++ETGKMPE     E+++ ALSYDE+K KM +E
Sbjct: 602 KLIAEALLKYETLDSTQIKALYETGKMPEAV--REESHALSYDEVEKMMNDE 651
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5165> which encodes the amino acid sequence <SEQ ID 5166>. Analysis of this protein sequence reveals the following:

```
        Possible site: 38

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL     Likelihood = -7.38    Transmembrane  138 - 154 ( 132 - 158)

----- Final Results -----
                bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC16243 GB:AF061748 cell division protein FtsH [Streptococcus pneumoniae] (ver 2)
 Identities = 487/654 (74%), Positives = 565/654 (85%), Gaps = 7/654 (1%)

Query: 5   KNNGFVKNSPIYILMITVVITGFQFYLKGTSTQ-SQQISYSKLIKHLKAGDIKSLGYQPS 63
           +NNG +KN P+++L I  ++TGFQ++  G ++  SQQI+Y++L++ +  G++K L+YQP+
Sbjct: 4   QNNGLIKNPFLWLLPIFFLVTGFQYFYSGNNSGGSQQINYYELVQEITDGNVKELTYQPN 63

Query: 64  GSIIEVKGKYKKPQKVTVNSGLSFLGGRASTQVTEPSSLVLPSDTILKEMTAAADKNSTE 123
           GS+IEV G Y+ P+      +G+ F       T+V +P+S +LP+DT + E+   A  + E
Sbjct: 64  GSVIEVSGVYKNPKTSKEGTGIQFPTPSV-TKVEKFTSTILPADTTVSELQKLATDHKAE 122

Query: 124 LTVKQESSSGTWITFLMSFLPIVIFAAFMMMMM-NQGGGGARGAMSFGKNKAKSQSKGNV 182
```

```
                +TVK ESSSG WI  L+S +P  I   F+  MM N GGG  R  MSPG++KAK+ +K ++
Sbjct: 123 VTVKHESSSGIWINLLVSIVPFGILFFFLFSMMGNMGGGNGRNFMSFGRSKAKAANKEDI 182

Query: 183 KVRFTDVAGAEEEKQELVEVVDFLKNPKKYKSLGARIPAGVLLEGPPGTGKTLLAKAVAG 242
           KVRF+DVAGAEEEKQELVEVV+FLK+PK++   LGARIPAGVLLEGPPGTGKTLLAKAVAG
Sbjct: 183 KVRFSDVAGAEEEKQELVEVVEFLKDPKRFTKLGARIPAGVLLEGPPGTGKTLLAKAVAG 242

Query: 243 EAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAERAIIFIDEIDAVGRRRGAGMGGG 302
           EAGVPFFSISGSDFVEMFVGVGASRVRSLFKDAKKA  AIIFIDEIDAVGR+RG G+GGG
Sbjct: 243 EAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAAPAIIFIDEIDAVGRQRGVGLGGG 302

Query: 303 NDEREQTLNQLLIEMDGFEGNENIIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDVKGR 362
           NDEREQTLNQLLIEMDGFEGNE IIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDVKGR
Sbjct: 303 NDEREQTLNQLLIEMDGFEGNEGIIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDVKGR 362

Query: 363 EAILRVHAKNKPLAEDVDLKLVAQQTPGFVGADLENVLNEAALVAARRNKIKIDASDIDE 422
           EAIL+VHAKNKPLA DV+LK+VAQQTPGFVGADLENVLNEAALVAARRNK  IDASDIDE
Sbjct: 363 EAILKVHAKNKPLAEDVDLKLVAQQTPGFVGADLENVLNEAALVAARRNKSIIDASDIDE 422

Query: 423 AEDRVIAGPSKKDRTISQKERERMVAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGGYMI 482
           AEDRVIAGPSKKD+T+SQKERE+VAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGGYMI
Sbjct: 423 AEDRVIAGPSKKDKTVSQKERELVAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGGYMI 482

Query: 483 ALPKEDQMLLSKEDLKEQLAGLMGGRVAEEIVFNAQTSGASNDFEQATQIARAMVTEYGM 542
           ALPKEDQMLLSKED+KEQLAGLMGGRVAEEI+FN QT+GASNDFEQATQ+ARAMVTEYGM
Sbjct: 483 ALPKEDQMLLSKEDMKEQLAGLMGGRVAEEIIFNVQTTGASNDFEQATQMARAMVTEYGM 542

Query: 543 SEKLGPVQYEGNHAMMPGQISPEKAYSAQTAQMIDDEVRELLNQARNQAADIINENRDTH 602
           SEKLGPVQYEGNHAML GAQ- SPQKSISEQTAYEIDEEVRSLLNEARNKAAEIIQSNRETH 601
Sbjct: 543 SEKLGPVQYEGNHAMLGAQ-SPQKSISEQTAYEIDEEVRSLLNEARNKAAEIIQSNRETH 601

Query: 603 KLIAEALLKYETLDAAQIKSIYETGKMPVDLSTDDNEAHALSYDEIKNKMTESE 656
           KLIAEALLKYETLD+ QIK++YETGKMP   E  + E+HALSYDE+K+KM + +
Sbjct: 602 KLIAEALLKYETLDSTQIKALYETGKMP---EAVEESSHALSYDEVKSKMNDEK 652
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 550/657 (83%), Positives = 612/657 (92%), Gaps = 2/657 (0%)

Query: 1   MKNNKNNGFKNSFIYILLIIAVITTFQYYLKGTSSQNQQISYTKLVKQLKAGEIKSISY 60
           MKNNKNNGF+KNSFIYIL+II  VITTFQ+YLKGTS+Q+QQISY+KL+K LKAG+IKS+SY
Sbjct: 1   MKNNKNNGFQKNSFIYILMIIVVITTGFQFYLKGTSTQSQQISYSKLIKHLKAGDIKSLSY 60

Query: 61  QPSGGVVEVSGTYKKAKTIKSANSFTFLGSSVATKVTGPNSVILPNDSSIKSLVSAAEEM 120
           QPSG ++EV G Y+K + +   + +FLGG  +T+VT P+S++LP+D+ +K + +AA++N
Sbjct: 61  QPSGSIIEVKGKYEKPQKVTVNSGLSPLGGRASTQVTEPSSLVLPSDTTLKEMTAAADKN 120

Query: 121 NTNIQVKIESSSGTWISYIASFLPLVIMIGFWMMMNQGGGGARGAMSPGKNKARSSSK 180
            T + VK ESSSGTWI+++ SFLP+VI   F MMMMNQGGGG ARGAMSPGKNKA+S SK
Sbjct: 121 GTELTVKQESSSGTWITFLMSFLPIVIFAAFMMMMMNQGGGG-ARGAMSPGKNKAKSQSK 179

Query: 181 DEVKVRFSDVAGAEEEKQELIKVVDFLKDPKRYKSLGARIPAGVLLEGPPGTGKTLLAKA 240
             VKVRF+DVAGAEEEKQEL+EVVDFLK+PK+YKSLGARIPAGVLLEGPPGTGKTLLAKA
Sbjct: 180 GNVKVRFTDVAGAEEEKQELVEVVDFLKNPKKYKSLGARIPAGVLLEGPPGTGKTLLAKA 239

Query: 241 VAGEAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAERAIIFIDEIDAVGRRRGAGM 300
           VAGEAGVPFFSISGSDFVEMFVGVGASRVRSLFEDAKKAERAIIFIDEIDAVGRRRGAGM
Sbjct: 240 VAGEAGVPFFSISGSDFVRMFVGVGASRVRSLFEDAKKAERAIIFIDEIDAVGRRRGAGM 299

Query: 301 GGGNDEREQTLNQLLIEMDGFEGNESIIVIAATNRSDVLDPALLRPGRFDRKVLVGQPDV 360
           GGGNDEREQTLNQLLIEMDGFEGNE+IIVIAATNRSDVLDPALLRPGRFDRKVLVG+PDV
Sbjct: 300 GGGNDEREQTLNQLLIEMDGFEGNENIIVIAATNRSDVLDPALLRPGRFDRKVLVGRPDV 359

Query: 361 KGREAILRVHAKNKPLADNVDLKVVAQQTPGFVGADLENVLNEAALVAARRNKKVIDASD 420
           KGREAILRVHAKNKPLA++V+LKVVAQQTPGFVGADLENVLNEAALVAARRNK  IDASD
Sbjct: 360 KGREAILRVHAKNKPLANDVNLKVVAQQTPGFVGADLENVLNEAALVAARRNKIKIDASD 419

Query: 421 IDEAEDRVIAGPSKKDRTISERERAMVAYHRAGHTIVGLILSNARVVHKVTIVPRGRAGG 480
```

```
              IDEAEDRVIAGPSKKDRTIS++ER MVAYHEAGHTIVGL+LSNARVVHKVTIVPRGRAGG
Sbjct: 420 IDEAEDRVIAGPSKKDRTISQKRREMVAYHEAGHTIVGLVLSNARVVHKVTIVPRGRAGG 479

Query: 481 YMIALPKRDQMLLSKDDMKEQLAGLMGGRVAEKIIFNAQTTGASNDFEQATAMARAMVTE 540
           YMIALPKRDQMLLSK+D+KEQLAGLMGGRVAEKI+FNAQT+GASNDFEQAT +ARAMVTE
Sbjct: 480 YMIALPKRDQMLLSKEDLKEQLAGLMGGRVAEKIVFNAQTSGASNDFEQATQIARAMVTE 539

Query: 541 YGMSEKLGPVQYEGNHAMMAGQMSPEKSYSAQTAQLIDDEVRHLLNRARMKAADIINENR 600
           YGMSEKLGPVQYEGNHAMM GQ+SPEK+YSAQTAQ+IDDEVR LLN+ARN+AADIINENR
Sbjct: 540 YGMSEKLGPVQYEGNHAMMPGQISPEKAYSAQTAQMIDDEVRELLNQARNQAADIINENR 599

Query: 601 DTHKLIAEALLKYETLDAAQIKSIFETGKMP-ETENDEDKARALSYDEIKEKMQEED 656
           DTHKLIAEALLKYETLDAAQIKSI+ETGKMP + E D+++A ALSYDEIK KM E +
Sbjct: 600 DTHKLIAEALLKYETLDAAQIKSIYETGKMPVDLETDDNEAHALSYDEIKNKMTESE 656
```

SEQ ID 5164 (GBS115) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 8; MW 73kDa) and in Figure 39 (lane 3; MW 73.3kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1666

A DNA sequence (GBSx1769) was identified in *S.agalactiae* <SEQ ID 5167> which encodes the amino acid sequence <SEQ ID 5168>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2983(Affirmative) < succ>
               bacterial membrene --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1667

A DNA sequence (GBSx1770) was identified in *S.agalactiae* <SEQ ID 5169> which encodes the amino acid sequence <SEQ ID 5170>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2424(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9547> which encodes amino acid sequence <SEQ ID 9548> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12187 GB:Z99106 similar to homoserine dehydrogenase [Bacillus subtilis]
 Identities = 223/448 (49%), Positives = 313/448 (69%)
```

```
Query: 1    MKVVKPGGSSLASGQQLYKVLNIIKSDYTRRPVVVSAPGKRYEEDLKMTDALIQYYQNYI 60
            MKVVKPGGSSLAS  QL KV +I+ SD  R+ VVVSAPGK Y ED K+TD LI   + Y+
Sbjct: 1    MKVVKPGGSSLASGAQLDKVSHIVTSDPARKAVVVSAPGKHYAEDTKVTDLLIACAEQYL 60

Query: 61   NGKDIVKDQTWIINRVQEIISDLSLGSTIAEETRSIEQLASLPIENNQFLYDCFLAAGE 120
                    +    ++ RY  I ++L LG +I E+I   +  L     N +  D   A+GE
Sbjct: 61   ATGSAPELAERAVVERYALIANELQLGQSIIEKIRDDLFTLLEGDKSNPEQYLDAVKASGE 120

Query: 121  DNNAKLVATFFNQNDIPARVVHPNEAGIIVTKEPCNARIIPGSYDKIENNLCLYNEVLWIP 180
            DNNAKL+A  +F    + A YV+P +AG+ VT EP NA+++P SY  +  L  + +++ P
Sbjct: 121  DNNAKLIAAYFRYKGVKAEYVNPKDAGLFVTNEPGNAQVLPESYQNLYRLRERDGLIIFP 180

Query: 181  GFFGVTEDMQICTFSRGSGSDITGSLIAAGIKADLYENFTDVDGIPAAHPGVVKNPHAIPE 240
            GFFG ++D  + TFSR GSDITGS++A G++ADLYENFTDVD +++ +P  V+NP  I E
Sbjct: 181  GFFGFSKDGDVITFSRGSGSDITGSLIANGLQADLYENFTDVDAVYSVNPSFVENPKSISE 240

Query: 241  LTYKEMRELAYAGFSVLHDEALLPAYRGRIPLVIKNTNNDQQPGTKIVLKHTRSNIAVTG 300
            LTY+EMRELAYAGFSV HDEALPAFRAGIPVQIKNTNNPSAEGTRVVSKRDNTNGPVVG 300
Sbjct: 241  LTYREMRELSYAGFSVFHDEALIPAFRAGIPVQIKNTNNPSAEGTRVVSKRDNTNGPVVG 300

Query: 301  IASDSIFASINVSKVLMNREWGFGEKVLQILEDLNISFEHMPTGIDDLSIVLREKELTPI 360
            IASD+ F SI +SKVLMNREWGFGR+ LQILE+  +++EH+P+GIDD++I+LR+ ++
Sbjct: 301  IASDTGFCSIYISKVLMNREIGFGRRALQILEEHGLTYEHVPSGIDDMTILRQGQMDAA 360

Query: 361  KEQSILNYLTRKLEVDYVDIQHNLSTIVIVGENMKSQIGVTATTATQALSREKINTIMISQ 420
              E+ ++  +   L  D V ++H+L+ I++VGE M+  +G TA A +ALS  ++NI MI+Q
Sbjct: 361  TERSVIKRIEEDLHADEVIVEHHLALINVVGEAMRHNVGTTARAAKALSEAQVNIEMINQ 420

Query: 421  GSSEVSIMFVINSKDEKRAIKALYETFF 448
            GSSEVS+MF +   +E++A++ALY+ FF
Sbjct: 421  GSSEVSMMFGVKEAEERKAVQALYQEFF 448
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1668

A DNA sequence (GBSx1771) was identified in *S.agalactiae* <SEQ ID 5171> which encodes the amino acid sequence <SEQ ID 5172>. This protein is predicted to be CbbY family protein. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
        bacterial cytoplasm --- Certainty=0.2699(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF96016 GB:AE004353 CbbY family protein [Vibrio cholerae]
 Identities = 59/190 (31%), Positives = 93/190 (48%), Gaps = 10/190 (5%)

Query: 4    YKAIIFDMDGVLPDTELFYYKRERRFPLKQHGITIDHLPMNFFIGGNMKQVWKSVLGDCYD 63
             ++A IFDMDG+L DTE    +  +     G+          IG N K +   +L   Y
Sbjct: 6    FQAAIFDMDGLLLDTERVCMRVFQRACTACGLPFRQEVYLSVIGCNAKTI-NGILSQAYG 64

Query: 64   TWDIDKL----QQDYSRYKEDNPLPYKDLIPQDCKRVIEKLHHKGYLLGLASSSTRHDIM 119
               D+ +L    +Q Y+        +P+KD +      ++E L  +   + +A+S+ +   +
Sbjct: 65   E-DLPRLHNEWRQRYNAVVMHRAIPHKDGVIA----LLEWLKARSIPVAVATSTQKEVAL 119

Query: 120  LALESFNLDTYFKVILSGEEFSESKPNPAIYNRAAELLDIPKQQILIVEDSEKGITAGIA 179
```

```
            + L+   LD YF  I +G E ++ KP+P IY  AAR L +  QQ L  EDS  GI A +A
Sbjct: 120 IKLQLAGLDHYFANITTGCEVTQGKPHPEIYLLAAERLGVSPQQCLAFEDSNNGIKAAMA 179

Query: 180 AGIDVWAIED 189
           A +  + I D
Sbjct: 180 AQMHAFQIPD 189
```

There is also homology to SEQ ID 448.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1669

A DNA sequence (GBSx1772) was identified in *S.agalactiae* <SEQ ID 5173> which encodes the amino acid sequence <SEQ ID 5174>. This protein is predicted to be Pseudomonas putida enoyl-CoA hydratase II homologue (b1394). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.18   Transmembrane  128 - 144 ( 128 - 145)
   INTEGRAL    Likelihood = -1.06   Transmembrane  154 - 170 ( 154 - 170)

----- Final Results -----
                bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9549> which encodes amino acid sequence <SEQ ID 9550> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5175> which encodes the amino acid sequence <SEQ ID 5176>. Analysis of this protein sequence reveals the following:

```
     Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.08   Transmembrane  110 - 126 ( 109 - 128)

----- Final Results -----
                bacterial membrane --- Certainty=0.2232(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 150/263 (57%), Positives = 197/263 (74%)

Query: 19  LKFENIYYGIDGNVATIMLNRPDISNGFNIPMCQEIIDAIRLVSENKDVMFLVIEAQGPI 78
           ++F++II+ +  ++AT+ LNRP++SNGFNIP+CQEI+ A+  V  +  V FL+I+A G +
Sbjct: 1   MQFKHIIPDVVDLATLTLNRPEVSNGFNIPICQEILVALAEVKRDTSVRFLLIKAVGKV 60

Query: 79  FSIGGDLKVMGAAVESDISSLTKIAELVNQISYDLLQLEKPVVMCVDGAVAGAAANIAL 138
           FS+GGDL  M+ AV  D++  SL KIAELV +IS+ +  L KPV++C DGAVAGAA NIAL
Sbjct: 61  FSVGGDLVEMQEAVAKDNVQSLVKIAELVQEISFAIKHLPKPVILCADGAVAGAAFNIAL 120

Query: 139 AADFVIASKKSKFIQAFVGVGLAPDAGGLLLLSKSIGITRAVQLALTGESLSAEKAEALG 198
           A DF IAS ++KFIQAFV VGLAPDAGGL LL+++G+ RA  L +TGE ++A+K   G
Sbjct: 121 AVDFCIASTQTKFIQAFVNVGLAPDAGGLFLLTRAVGLNRATHLVMTGEGITADKGLDYG 180

Query: 199 IVYKLCESDKIGKIKDQLLKRLSRHSINSYQAIKSLAWEAAFKDWEQYKKLELQLQESLA 258
            VY+  ESDK+ K+  QLLKRL R S NSY  +KSL W++ F  WE Y K EL +QE LA
Sbjct: 181 FVYRTAESDKLDKVCLQLLKRLARGSSNSYAGMKSLVWQSFFTGWKDYAKALAIQESLA 240
```

```
Query: 259 FKQDFKEGVRAHADRRRPNFLGK 281
          FK+DFKEGV A  +RRRPNF GK
Sbjct: 241 FKEDFKEGVIAFGERRRPNFQGK 263
```

A related GBS gene <SEQ ID 8877> and protein <SEQ ID 8878> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
SRCFLG: 0
McG: Length of UR:   9
     Peak Value of UR:   1.45
     Net Charge of CR: -1
McG: Discrim Score:     -5.99
GvH: Signal Score (-7.5): -4.37
     Possible site: 27
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program   count: 2 value:  -2.18 threshold:  0.0
   INTEGRAL   Likelihood = -2.18   Transmembrane  110 - 126 ( 110 - 127)
   INTEGRAL   Likelihood = -1.06   Transmembrane  136 - 152 ( 136 - 152)
   PERIPHERAL  Likelihood =  1.32      49
 modified ALOM score:   0.94
icm1 HYPID: 7  CFP: 0.187

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01047(355 - 1143 of 1443)
GP|3253198|gb|AAC24330.1||AF029714(1 - 263 of 263) PhaB {Pseudomonas putida}
%Match = 15.4
%Identity = 33.3  %Similarity = 56.4
Matches = 88  Mismatches = 113  Conservative Sub.s = 61

96        126        156        186        216        246        276        306
*KTVRRGLQLVLQFVLMCGLLKINTLE*ISRRLMY**AI*VNFL*N*ITTKNGKFNSVFLFFILP*KLGL**NTKHDNLI

336        366        396        426        456        486        516        546
IKLFFIFLSLLKRGDILKFENIIYGIDGNVATIMLNRPDISNGFNIPMCQEIIDAIRLVSENKDVMFLVIEAQGDIFSIG
                    MTFQHILFSIEDGVAFLSLNRPEQLNSFNAAMHLEVREALKQVRQSSDARVLLLTAEGRGFCAG
                              10         20         30         40         50         60

576        606        636        666        696        726        756        786
GDLKVMKAAVESDDISSLTKIAELVNQISYDLLQLEKPVVMCVDGAVAGAAANIALAADFVIASKKSKFIQAFVGVGLAP
QDLSDRNVAPDAEVPDLGESIDKFYNPLVKTLRDLPLPVICAVNGVAAGAGANIPLACDLVLAGRSASFIQAFCKIGLVP
                    80         90         100        110        120        130        140

816        846        876        906        936        966        996        1026
DAGGLLLLSKSIGITRAVQLALTGESLSAEKAEALGIVVKLCESDKIGKIKDQLLKRLSRHSINSYQAIKSLAWEAAFKD
DSGGTWLLPRLVGMARAKALAMLGERLGABQAQQWGLIHRVVDDAALRDEALTLARQLASQPTYGLALIK-RSLNASFDN
                    160        170        180        190        200        210        220

1053       1083       1113       1143       1173       1203       1233       1263
-WEQYKKLELQLQSSLAFKQDFKEGVRAHADRRRPNFLGK*FENQII*D*SLANKFEL*YNLIIKV*CEVVISWNTIRLI
GFDEQLELERDLQRLAGRSEDYREGVSAFMNKRTPAFKGR
                    240        250        260
```

SEQ ID 8878 (GBS374) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 8; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 71 (lane 2; MW 57kDa).

The GBS374-GST fusion product was purified (Figure 215, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 307), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1670

A DNA sequence (GBSx1773) was identified in *S.agalactiae* <SEQ ID 5177> which encodes the amino acid sequence <SEQ ID 5178>. This protein is predicted to be a 16.1 kDa transcriptional regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD05186 GB:AF110185 unknown [Burkholderia pseudomallei]
 Identities = 30/102 (29%), Positives = 60/102 (58%)

Query: 32  DVSLKEMHTIEIIGKHSEVTPSDVARELMLTLGTVTTSLNKLEKKGYIERKRSSIDRRVV 91
           +++ +++  I ++ +     TP +++R+L    G++T  L++LEKKG++ R RS  DRRV+
Sbjct: 39  ELTAQQISVILLLARGVARTPFELSRKLSYDSGSNTRMLDRLEKKGFVVRARSESDRRVI 98

Query: 92  HLSLTRRGRLLDRLHSKPHKSMVSHIIEDLGEEDIKMLTSAL 133
            L+LT+RG    R      + ++  +E    +++ +LT  L
Sbjct: 99  ELALTERGAHAARAADPALIATELNAQLEGFSADEVALLTDLL 140
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5179> which encodes the amino acid sequence <SEQ ID 5180>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1412(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 111/144 (77%), Positives = 129/144 (89%)

Query: 1   MEYDQINSYLVDIFNRIMIIEEMSLKTSQFSDVSLKEMHTIEIIGKHSEVTPSDVARELM 60
           +EYD+I  YLVDIFNRI++IEEMSLKTSQFSDVSLKEMHTIEIIGK+ +VTPSD+ARELM
Sbjct: 7   LEYDKIYPYLVDIFNRILVIEEMSLKTSQFSDVSLKEMHTIEIIGKYDQVTPSDIARELM 66

Query: 61  LTLGTVTTSLNKLEKKGYIERKRSSIDRRVVHLSLTRGRLLDRLHSKPHKSMVSHIIED 120
           +TLGTVTTSLNKLE KGYI R RS  DRRVV+LSLTRGRLLDRLH+KPHK+MV H+I D
Sbjct: 67  VTLGTVTTSLNKLEAKGYIARTRSRSDRRVVYLSLTRGRLLDRLHAKPHKNMVGHVIAD 126
```

```
Query: 121 LGEEDIKMLTSALGNLHKFLEDLV 144
           + +E+++ L   LGNLH+FLEDLV
Sbjct: 127 MSDEEMQALVRGLGNLHQFLEDLV 150
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1671

A DNA sequence (GBSx1774) was identified in *S.agalactiae* <SEQ ID 5181> which encodes the amino acid sequence <SEQ ID 5182>. This protein is predicted to be 3-oxoacyl-(acyl-carrier-protein) synthase III (fabH-2). Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.12   Transmembrane  103 - 119 ( 103 - 119)

----- Final Results -----
                bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98271 GB:AF197933 beta-ketoacyl-ACP synthase III
           [Streptococcus pneumoniae]
 Identities = 225/324 (69%), Positives = 276/324 (84%), Gaps = 1/324 (0%)

Query: 1   MVFAKISQLAHYAPSQIIKMEDLSLIMDTSDDWISSRTGIKQRHISKNETTADLANKVAE 60
           M FAKISQ+AHY P Q++ N DL+ IMDT+D+WISSRTGI+QRHIS+ E+T+DLA +VA+
Sbjct: 1   MAFAKISQVAHYVPRQVVTNHDLAQIMDTNDEWISSRTGIRQRHISRTESTSDLATEVAK 60

Query: 61  QLIEKSGYSASQIDFIIVATWTPDSMMPSTAARVQAHIGASNAFAFDLSAACSGFVFALS 120
           +L+ K+G +  ++DFII+AT+TPDSMMPSTAARVQA+IGA+ AFAFDL+AACSGFVFALS
Sbjct: 61  KLMAKAGITGEELDFIILATITPDSMMPSTAARVQANIGANKAFAFDLTAACSGFVFALS 120

Query: 121 TAEKLISSGSYQKGLVIGAETVSKVLDWTDRGTAVLFGDGAGGVLLEASKEKHFLAESLN 180
           TAEK I+SG +QKGLVIG+ETV+SK +DW+DR TAVLFGDGAGGVLLEAS+++HFLAESLN
Sbjct: 121 TAEKFIASGRPQKGLVIGSETLSKAVDWSDRSTAVLFGDGAGGVLLEASEQRHFLAESLN 180

Query: 181 TDGSR-QCLQSSQVGLNSPFSDEVLDDKFLKMDGRAIFDFAIKEVSKSINHLIETSYLRK 239
           +DGSR + L    GL+SPFSD+    D FLKMDGR +FDFAI++V+KSI   I+ S +E
Sbjct: 181 SDGSRSECLTYGHSGLHSPFSDQESADSFLKMDGRIVFDFAIRDVAKSIKQTIDESPIEV 240

Query: 240 EDIDYLFLHQANRRILDKMSRKIDIARDKFPENMMDYGNTSAASIPILLSESYENGLLKL 299
            D+DYL LHQAN RILDKM+RKI + R K P NMM+YGNTSAASIPILLSE  E GL+ L
Sbjct: 241 TDLDYLLLHQANRRILDKMARKIGVDRAKLPANMMEYGNTSAASIPILLSECVEQGLIPL 300

Query: 300 DGNQTILLSGFGGGLTWGSLIVKI 323
           DG+QT+LLSGFGGGLTWG+LI+ I
Sbjct: 301 DGSQTVLLSGFGGGLTWGTLILTI 324
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5183> which encodes the amino acid sequence <SEQ ID 5184>. Analysis of this protein sequence reveals the following:

```
      Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane  103 - 119 ( 103 - 120)

----- Final Results -----
                bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF98271 GB:AF197933 beta-ketoacyl-ACP synthase III
          [Streptococcus pneumoniae]
 Identities = 212/324 (65%), Positives = 263/324 (80%)

Query: 1   MIFSKISQVAHYVPQQLVTNNDLASIMDTSHEWIFSRTGIAERHISRDEMTSDLAIQVAD 60
            M F+KISQVAHYVP+Q+VTN+DLA IMDT+ EWI SRTGI +RHISR E TSDLA +VA
Sbjct: 1   MAFAKISQVAHYVPEQVVTNHDLAQIMDINDEWISSRTGIQRHISRTESTSDLATEVAK 60

Query: 61  QLLTQSGLKADAIDFIIVATISPDATMPSTAAKVQAAIAATSAFAFDMTAACSGFVFALA 120
            +L+ ++G+  + +DFII+ATI+PD+ MPSTAA+VQA I A  AFAFD+TAACSGFVFAL+
Sbjct: 61  KLMAKAGITGEELDFIILATITPDSMMPSTAARVQANIGANKAFAFDLTAACSGFVFALS 120

Query: 121 MADKLIASGAYQNGMVIGAETLSKLVNWQDRATAVLFGDGAGGVLLKASKDKHVLAETLH 180
             A+K IASG +Q G+VIG+ETLSK V+W DR+TAVLFGDGAGGVLLKAS+ +H IAR+L+
Sbjct: 121 TAEKFIASGRFQKGLVIGSETLSKAVDWSDRSTAVLFGDGAGGVLLKASEQEHFLAESLN 180

Query: 181 TDGARCQSLISGETSLSSPYSIGKKAIATIQMDGRAIFDFAIRDVSKSILTLMAQSDITK 240
            +DG+R + L  G + L SP+S  + A + ++MDGR +FDFAIRDV+KSI    + +S I
Sbjct: 181 SDGSRSECLTYGHSGLHSPFSDQESADSFLKMDGRTVFDFAIRDVAKSIKQTIDESPIEV 240

Query: 241 DDIDVCLLHQANRRILDKIARKIDVPREKFLENMMRYGNTSAASIPILLSEAVQKGQIRL 300
             D+DY LLHQAN RILDK+ARKI V R K   NMM YGNTSAASIPILLSE V++G I L
Sbjct: 241 TDLDYLLLHQANDRILDKMARKIGVDRAKLPANMMEYGNTSAASIPLLSECVEQGLIPL 300

Query: 301 DGTQKILLSGPGGGLTWGSLIVRI 324
            DG+Q +LLSGPGGGLTWG+LI+ I
Sbjct: 301 DGSQTVLLSGPGGGLTWGTLILTI 324
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 216/324 (66%), Positives = 271/324 (82%), Gaps = 1/324 (0%)

Query: 1   MVFAKISQLAHYAPSQIIYNEDLSLIMDTSDDWISSRTGIKQRHISKNETTADLANKVAE 60
            M+F+KISQ+AHY P Q++ N DL+ IMDTS +WI SRTGI +RHIS++E T+DLA +VA+
Sbjct: 1   MIFSKISQVAHYVPQQLVTNNDLASIMDTSHEWIFSRTGIAERHISRDEMTSDLAIQVAD 60

Query: 61  QLIEKSGYSASQIDFIIVATMTPDSMMPSTAARVQAHIGASNAFAFDLSAACSGFVFALS 120
            QL+ +SG  A  IDFIIVAT++PD+ MPSTAA+VQA I A++AFAFD++AACSGFVFAL+
Sbjct: 61  QLLTQSGLKADAIDFIIVATISPDATMPSTAAKVQAAIAATSAFAFDMTAACSGFVFALA 120

Query: 121 TAEKLISSGSYQKGLVIGAETVSKVLDWTDRGTAVLFGDGAGGVLLEASKEKHFLAESLN 180
             A+KLI+SG+YQ G+VIGAET+SK++W DR TAVLFGDGAGGVLLKASK+KH LAE+L+
Sbjct: 121 MADKLIASGAYQNGMVIGAETLSKLVNWQDRATAVLFGDGAGGVLLKASKDKHVLAETLH 180

Query: 181 TDGSR-QGLQSSQVGLNSPFSDEVLDDKFLKMDGRAIFDFAIKEVSKSINHLIETSYLEX 239
            TDG+R Q L S +  L+SP+S          ++MDGRAIFDFAI++VSKSI   L+  S + K
Sbjct: 181 TDGARCQSLISGETSLSSPYSIGKKAIATIQMDGRAIFDFAIRDVSKSILTLMAQSDITK 240

Query: 240 EDIDYLFLHQANRRILDKNSRKIDIARDKFPENMMDYGNTSAASIPILLSESYRNGLLKL 299
            +DIDY  LHQANRRILDK++RKID+ R+KF ENMM YGNTSAASIPILLSE+ + G ++L
Sbjct: 241 DDIDYCLLHQANRRILDKIARKIDVPREKFLENMMRYGNTSAASIPILLSEAVQKGQIRL 300

Query: 300 DGNQTILLSGPGGGLTWGSLIVKI 323
            DG Q ILLSGPGGGLTWGSLIV+I
Sbjct: 301 DGTQKILLSGPGGGLTWGSLIVRI 324
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1672

A DNA sequence (GBSx1775) was identified in *S.agalactiae* <SEQ ID 5185> which encodes the amino acid sequence <SEQ ID 5186>. This protein is predicted to be acyl carrier protein (acpP). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3083(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9551> which encodes amino acid sequence <SEQ ID 9552> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98272 GB:AF197933 acyl carrier protein [Streptococcus pneumoniae]
 Identities = 64/74 (86%), Positives = 67/74 (90%)

Query: 17 MAVFEKVQRIIVEELGKDAEEVTLNTTFDDLDADSLDVPQVISEIEDAFDIQIETEEGLN 76
          MAVFEKVQRIIVEELGKDA EVTL +TFDDLDADSLD+PQVISEIEDAFDIQIE E  L
Sbjct: 1  MAVFEKVQRIIVEELGKDASEVTLESTFDDLDADSLDLPQVISEIEDAFDIQIKAENDLK 60

Query: 77 TVGDLVAYVEEKVK 90
          TVGDLVAYVEE+ K
Sbjct: 61 TVGDLVAYVEEQAK 74
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5187> which encodes the amino acid sequence <SEQ ID 5188>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2995(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 70/74 (94%), Positives = 71/74 (95%)

Query: 17 MAVFEKVQRIIVEELGKDAEEVTLNTTFDDLDADSLDVPQVISEIEDAFDIQIETEEGLN 76
          MAVFEKVQRIIVEELGK+ EEVTL TTFDDLDADSLDVPQVISEIEDAFDIQIETEEGLN
Sbjct: 1  MAVFEKVQRIIVEELGKETEEVTLETTFDDLDADSLDVPQVISEIEDAFDIQIETEEGLN 60

Query: 77 TVGDLVAYVEEKVK 90
          TVGDLVAYVEEK K
Sbjct: 61 TVGDLVAYVEEKSK 74
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1673

A DNA sequence (GBSx1777) was identified in *S.agalactiae* <SEQ ID 5189> which encodes the amino acid sequence <SEQ ID 5190>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence
```

```
   INTEGRAL    Likelihood = -0.27   Transmembrane  156 - 172 ( 156 - 173)

  ----- Final Results -----
                bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98273 GB:AF197933 trans-2-enoyl-ACP reductase II
          [Streptococcus pneumoniae]
 Identities = 257/318 (80%), Positives = 277/318 (86%), Gaps = 1/318 (0%)

Query: 1   MKTRITELLNIKYPIFQGGMAWVADGDLAGAVSKAGGLGIIGGGNAPKEVVKANIDKIKS 60
           MKTRITBLL I YPIFQGGMAWVADGDLAGAVSKAGGLGIIGGGNAPKEVVKANIDKIKS
Sbjct: 1   MKTRITELLKIDYPIFQGGMAWVADGDLAGAVSKAGGLGIIGGGNAPKEVVKANIDKIKS 60

Query: 61  MTDKPFGVNIMLLSPFVDDIVDLVIEEGVKVVTTGAGNPGKYMERFHEAGITVIPVVPSV 120
           +TDKPFGVNIMLLSPFV+DIVDLVIEEGVKVVTTGAGNP KYMERFHEAGI VIPVVPSV
Sbjct: 61  LTDKPFGVNIMLLSPFVEDIVDLVIEEGVKVVTTGAGNPSKYMERFHEAGIVIPVVPSV 120

Query: 121 ALAKRMEKLGADAIITEGMEAGGHIGKLTTMTLVRQVVDAVTIPVIAAGGIADGRGAAAG 180
           ALAKRMEK+GADA+I EGMEAGGHIGKLTTMTLVRQV  A++IPVIAAGGIADG GAAAG
Sbjct: 121 ALAKRMEKIGADAVIAEGMEAGGHIGKLTTMTLVRQVATAISIPVIAAGGIADGEGAAAG 180

Query: 181 FMLGADAVQVGTRFVVAKESNAHPNYKAKILKAKDIDTAVSAQVVKGHPVRALKNKLVTTY 240
           FMLGA+AVQVGTRFVVAKESNAHPNYK KILKA+DIDT +SAQ  GH VRA+KN+L    +
Sbjct: 181 FMLGAEAVQVGTRFVVAKESNAHPNYKEKILKARDIDTTISAQHPGHAVRAIKNQLTRDF 240

Query: 241 SQAEKDYLAGRISINEI-EELGAGALRNAVVDGDVINGSVMAGQIAGLIKSEETCQEILE 299
             AEKD         EI E++GAGAL  AVV GDV  GSVMAGQIAGL+  EET +EIL+
Sbjct: 241 ELAEKDAFKQEDPDLEIFEQMGAGALAKAVVHGDVDGSVMAGQIAGLVSKEETAEEILK 300

Query: 300 DIYSGARQVILSEASRWS 317
           D+Y GA + I  EASRW+
Sbjct: 301 DLYYGAAKIQEEASRWT 318
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5191> which encodes the amino acid sequence <SEQ ID 5192>. Analysis of this protein sequence reveals the following:

```
      Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.70   Transmembrane  106 - 122 ( 106 - 124)
   INTEGRAL    Likelihood = -0.22   Transmembrane  156 - 172 ( 156 - 173)

  ----- Final Results -----
                bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF98273 GB:AF197933 trans-2-enoyl-ACP reductase II
          [Streptococcus pneumoniae]
 Identities = 252/320 (78%), Positives = 276/320 (85%), Gaps = 1/320 (0%)

Query: 1   MKTRITELLNIDYPIFQGGMAWVADGDLAGAVSNAGGLGIIGGGNAPKEVVKANIDRVKA 60
           MKTRITELL IDYPIFQGGMAWVADGDLAGAVS AGGLGIIGGGNAPKEVVKANID++K+
Sbjct: 1   MKTRITELLKIDYPIFQGGMAWVADGDLAGAVSKAGGLGIIGGGNAPKEVVKANIDKIKS 60

Query: 61  ITDPPGVNIMLLSPFNDDIVDLVIEEGVKVVTTGAGNPGKYMERLHQAGIIVVPVVPSV 120
           +TD+PPGVNIMLLSPP +DIVDLVIEEGVKVVTTGAGNP KYMER H+AGIIV+PVVPSV
Sbjct: 61  LTDKPFGVNIMLLSPFVEDIVDLVIEEGVKVVTTGAGNPSKYMERFHEAGIIVIPVVPSV 120

Query: 121 ALAKRMEKLGVDAVIABGMEAGGHIGKLTTMSLVRQVVEAVSIPVIAAGGIADGHGAAAA 180
           ALAKRMEK+G DAVIARGMEAGGHIGKLTTM+LVRQV  A+SIPVIAAGGIADG GAAA
Sbjct: 121 ALAKRMEKIGADAVIARGMEAGGHIGKLTTMTLVRQVATAISIPVIAAGGIADGHGAAAG 180
```

```
Query: 181 FMLGAEAVQIGTRFVVAKESNAHQNFKDKILAAKDIDTVISAQVVGHPVRSIKNKLTSAY 240
           FMLGAEAVQ+GTRFVVAKESNAH N+K+KIL A+DIDT ISAQ  GH VR+IKN+LT  +
Sbjct: 181 FMLGAEAVQVGTRFVVAKESNAHPNYKEKILKARDIDTTISAQHPGHAVRAIKNQLTRDF 240

Query: 241 AKAEK-AFLIGQKTATDIEEMGAGSLRHAVIEGDVVNGSVMAGQIAGLVRKEESCETILK 299
            AEK AF          E+MGAG+L  AV+ GDV  GSVMAGQIAGLV KEE+ E ILK
Sbjct: 241 ELAEKDAPKQEDPDLEIPSQMGAGALAKAVVHGDVDGGSVMAGQIAGLVSKESTAEEILK 300

Query: 300 DIYYGAARVIQNEAKRWQSV 319
           D+YYGAA+ IQ EA RW  V
Sbjct: 301 DLYYGAAKIQEEASRWTGV 320
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 253/319 (79%), Positives = 291/319 (90%)

Query: 1   MKTRITELLNIKYPIPQGGMAWVADGDLAGAVSKAGGLGIIGGGNAPKEVVKANIDKIKS 60
           MKTRITELLNI YPIPQGGMAWVADGDLAGAVS AGGLGIIGGGNAPKEVVKANIDK+K+
Sbjct: 1   MKTRITELLNIDYPIPQGGMAWVADGDLAGAVSNAGGLGIIGGGNAPKEVVKANIDRVKA 60

Query: 61  MTDKPFGVNIMLLSPFVDDIVDLVIEEGVKVVTTGAGNPGKYMERFHEAGITVIPVVPSV 120
           +TD+PFGVNIMLLSPF DDIVDLVIEEGVKVVTTGAGNPGKYMER H+AGI V+PVVPSV
Sbjct: 61  ITDRPFGVNIMLLSPFADDIVDLVIEEGVKVVTTGAGNPGKYMERLHQAGIIVVPVVPSV 120

Query: 121 ALAKRMEKLGADAIITEGMEAGGHIGKLTMTVLVRQVVDAVTIPVIAAGGIADGRGAAAG 180
           ALAKRMEKLG DA+I EGMEAGGHIGKLTM+LVRQVV+AV+IPVIAAGGIADG GAAA
Sbjct: 121 ALAKRMEKLGVDAVIAEGMEAGGHIGKLTMSLVRQVVEAVSIPVIAAGGIADGHGAAAA 180

Query: 181 FMLGADAVQVGTRFVVAKESNAHPNYKAKILKAKDIDTAVSAQVVGHPVRALKNKLVTTY 240
           FMLGA+AVQ+GTRFVVAKESNAH N+K KIL AKDIDT +SAQVVGHPVR++KNKL + Y
Sbjct: 181 FMLGAEAVQIGTRFVVAKESNAHQNFKDKILAAKDIDTVISAQVVGHPVRSIKNKLTSAY 240

Query: 241 SQAEKDYLAGRISINEIEELGAQALRNAVVDGDVINGSVMAGQIAGLIKSEETCQEILED 300
           ++AEK +L G+ +  +IEE+GAG+LR+AV++GDV+NGSVMAGQIAGL++ EE+C+ IL+D
Sbjct: 241 AKAEKAFLIGQKTATDIEEMGAGSLRHAVIEGDVVNGSVMAGQIAGLVRKEESCETILKD 300

Query: 301 IYSGARQVILSEASKWSDL 319
           IY GA +VI +EA RW  +
Sbjct: 301 IYYGAARVIQNEAKRWQSV 319
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1674

A DNA sequence (GBSx1778) was identified in *S.agalactiae* <SEQ ID 5193> which encodes the amino acid sequence <SEQ ID 5194>. This protein is predicted to be MCAT (fabD). Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1276(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with a *S.pneumoniae* sequence:

```
Identities = 203/306 (66%), Positives = 242/306 (78%), Gaps = 1/306 (0%)

Query: 1   MNKVSFLFAGQGAQKLGMARDLYETFPIVKETFDKASHVLGYDLRELIDKDLDKLNQTKY 60
           M K +FLFAGQGAQ LGM RD Y+ +PIVKET D+AS VLGYDLR LID + DKLNQT+Y
Sbjct: 1   MIKTAFLFAGQGAQYLGMGRDFYDQYPIVKETIDRASQVLGYDLRYLIDTEEDKLNQTRY 60
```

```
Query: 61  TQPAILTTSTAIYRLILKEIELRPDMVAGLSLGEYSALVASGAIRFEDAVVLVARRGQLM 120
           TQPAIL TS AIYRL L+E   +PDMVAGLSLGEYSALVASGA+ FEDAV LVA+RG  M
Sbjct: 61  TQPAILATSVAIYRLL-LQEKGYQPDMVAGLSLGEYSALVASGALDFEDAVALVARRGAYM 119

Query: 121 EAAAPAGSGKMVAVLNADRQIIEDACKKASQFGIVSPANYNTPKQIVIGGESIAVNAAVE 180
           E AAPA SGKMVAVLN   ++IE+AC+KAS+ G+V+PANYNTP QIVI GE +AV+ AVE
Sbjct: 120 EEAAPADSGKMVAVLNTPVEVIEEACQKASELGVVTPANYNTPAQIVIAGEVVAVDRAVE 179

Query: 181 ELKQQGVKRLIPLNVSGPFHTALLKPASQKLSDVLDKVHFSVSEIPVIGNTEAQIMKDDD 240
           L++ G KRLIPL VSGPFHTALL+PASQKL++ L +V FS    P++GNTEA +M+K+D
Sbjct: 180 LLQSAGAKRLIPLKVSGPFHTALLEPASQKLAETLAQVSFSDFTCPLVGNTEAAVMQKED 239

Query: 241 IKSLLARQVMEPVRFDESIETMKKMGMTQVVEIGPGKVLSGFLKKIDSSLSVHSVEDKIQ 300
           I  LL RQV EPVRF ESI  M++ G++  +EIGPGKVLSGF+KKID + +  VED+
Sbjct: 240 IAQLLTRQVKEPVRFYESIGVMQRAGISNFIEIGPGKVLSGFVKKIDQTAHLAHVEDQAS 299

Query: 301 FNNLKE 306
           L E
Sbjct: 300 LVALLE 305
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5195> which encodes the amino acid sequence <SEQ ID 5196>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1602(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 201/299 (67%), Positives = 248/299 (82%), Gaps = 1/299 (0%)

Query: 1   MNKVSFLFAGQGAQKLGMARDLYETFPIVKETFDKASHVLGYDLRELIDKDLDKLNQTKY 60
            M K +FLFAGQGAQKLGMARD Y+ F IV++TFD+AS VLGYDLR LID D  KLNQT Y
Sbjct: 3   MTKTAFLFAGQGAQKLGMARDFYDNFAIVKETFDQASQVLGYDLRRLIDSDELKLNQTSY 62

Query: 61  TQPAILTTSTAIYRLILKEIELRPDMVAGLSLGEYSALVASGALSFEDTLSLVARKGRLM 120
            TQPAILT+S AIYR +L     ++PDMVAGLSLGEYSALVASGA+ FED + LVA+RG+LM
Sbjct: 63  TQPAILTSSIAIYR-VLGLHHVKPDMVAGLSLGEYSALVASGALSFEDTLSLVAKRGRLM 121

Query: 121 EAAADAGSGKMVAVLNADRQIIEDACKKASQFGIVSPANYNTPKQIVIGGESIAVNAAVE 180
            E AAP GSGKMVAV+N D Q+IE+ C+ A++ G+V+PANYNTP QIVIGG++ AVN AVE
Sbjct: 122 EEAAPQGSGKMVAVMNTDVQVIEEVCQIAAKHGVVAPANYNTPSQIVIGGQTDAVNVAVE 181

Query: 181 ELKQQGVKRLIPLNVSGPFHTALLKPASQKLSDVLDKVHFSVSEIPVIGNTEAQIMKDDD 240
            LK++GVKRLIPLNVSGPFHTALL+PAS+ L+  L++ +FS  +IP++GNTEA IM+KD
Sbjct: 182 LLKERGVKRLIPLNVSGPFHTALLEPASRLLAKELERYNFSDPKIPLVGNTEANIMEKDR 241

Query: 241 IKSLLARQVMEPVRFDESIETMKKMGMTQVVEIGPGKVLSGFLKKIDSSLSVHSVEDKI 299
            I  LLARQVMEPVRF +S+ T+ + G+TQ +E+GPGKVL+GF+KKID KNLLCTSVEMMV 300
Sbjct: 242 IPELLARQVMEPVRFYDSVATLVESGITQPIEVGPGKVLTGFVKKIDKNLLCTSVEMMV 300
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1675

A DNA sequence (GBSx1779) was identified in *S.agalactiae* <SEQ ID 5197> which encodes the amino acid sequence <SEQ ID 5198>. This protein is predicted to be beta-ketoacyl-ACP reductase (fabG). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0930(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98275 GB:AF197933 beta-ketoacyl-ACP reductase [Streptococcus pneumoniae]
 Identities = 184/243 (75%), Positives = 212/243 (86%)

Query: 1   MQLKDKNIFITGSSRGIGLAIAHQFAQLGANIVLNGRSRISRDLIAEFADVGVKVIAISG 60
           M+L+ KNIFITGSSRGIGLAIAH+FAQ GANIVLN R  ISE+L+AEF++YG+KV+ ISG
Sbjct: 1   MKLEHKNIFITGSSRGIGLAIAHKFAQAGANIVLNSRGAISEELLAEFSNYGIKVVPISG 60

Query: 61  DVSSFEDANRMIKEAIASLGSVDVLVNNAGITNDKLMLKMTVEDFESVLKINLTGAFNMT 120
           DVS F DA RMI +AIA LGSVDVLVNNAGIT D LMLKMT  DFE VLK+NLTGAFNMT
Sbjct: 61  DVSDFADAKRMIDQAIAELGSVDVLVNNAGITQDTLMLKMTEADFEKVLKVNLTGAFNMT 120

Query: 121 QSVLKPMTKARQGAIINISSVVGLTGNVGQANYAASKAGLIGFTKSVAREVAARGIRVNA 180
           QSVLKPM KAR+GAIIN+SSVVGL GN+GQANYAASKAGLIGFTKSVAREVA+R IRVN
Sbjct: 121 QSVLKPMMKARRGAIINMSSVVGLMGNIGQANYAASKAGLIGFTKSVAREVASRNIRVNV 180

Query: 181 IAPGFIESDMTDVIPEKMQEAILAQIPMKRIGKKEVAQVASFLASQEYLTGQVIAIDGG 240
           IAPG IESDMT ++ +K++EA LAQIPMK  G+ ++VA +  FLA Q+YLTGQV+AIDGG
Sbjct: 181 IAPGMIESDMTAILSDKIKEATLAQIPMKEPGQAEQVADLTVFLASQDYLTGQVVAIDGG 240

Query: 241 MTM 243
           ++M
Sbjct: 241 LSM 243
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3865> which encodes the amino acid sequence <SEQ ID 3866>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1088(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 201/244 (82%), Positives = 220/244 (89%)

Query: 1   MQLKDKNIFITGSSRGIGLAIAHQFAQLGANIVLNGRSRISRDLIAEFADVGVKVIAISG 60
           M++K KNIFITGS+RGIGLA+AHQFA L ANIVLNGRS ISE+L+A F DYGV V+ ISG
Sbjct: 1   MEIKGKNIFITGSTRGIGLAMAHQFASLEANIVLNGRSAISKELVASFTDYGVTVVTISG 60

Query: 61  DVSSFEDANRMIKEAIASLGSVDVLVNNAGITNDKLMLKMTVKDFESVLKINLTGAFNMT 120
           DVS   +A RM+ EAI SLGS+DVLVNNAGITNDKLMLKMT EDFE VLKINLTGAFNMT
Sbjct: 61  DVSEASSAKRMVNEAIESLGSIDVLVNNAGITNDKLMLKMTKEDFERVLKINLTGAFNMT 120

Query: 121 QSVLKPMTKARQGAIINISSVVGLTGNVGQANYAASKAGLIGFTKSVAREVAARGIRVNA 180
           QSVLKPM KARQGAIIN+SSVVGLTGN+GQANYAASKAG+IGFTKSVAREVAAR I VNA
Sbjct: 121 QSVLKPMIKARQGAIINVSSVVGLTGNIGQANYAASKAGMIGFTKSVAREVAARNICVNA 180

Query: 181 IAPGFIESDMTDVIPEKMQEAILAQIPMKRIGKKEVAQVASFLASQEYLTGQVIAIDGG 240
           IAPGFIESDMT V+PEKMQE IL+QIPMKRIGK +EVA +ASFL SQ+Y+TGQVIAIDGG
Sbjct: 181 IAPGFIESDMTGVLPEKMQEQILSQIPMKRIGKAQEVAHLASFLVSQDYITGQVIAIDGG 240

Query: 241 MTMQ 244
           MTMQ
Sbjct: 241 MTMQ 244
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1676

A DNA sequence (GBSx1780) was identified in *S.agalactiae* <SEQ ID 5199> which encodes the amino acid sequence <SEQ ID 5200>. This protein is predicted to be 3-oxoacyl-(acyl-carrier-protein) synthase II (fabF). Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane  338 - 354 ( 338 - 354)

----- Final Results -----
                bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAP98276 GB:AF197933 beta-ketoacyl-ACP synthase II
           [Streptococcus pneumoniae]
 Identities = 340/410 (82%), Positives = 375/410 (90%)

Query: 1   MTLQRVVVTGYGVTSPIGNTPEEFWNSLKEGNVGIGPITKFDSSDFMVKNAAEIHDFPFD 60
            M L RVVVTGYGVTSPIGNTPEEFWNSL  G +GIG ITKFD SDF V NAAEI DFPFD
Sbjct: 1   MKLNRVVVTGYGVTSPIGNTPEEFWNSLATGKIGIGGITKFDHSDFDVHNAAEIQDFPFD 60

Query: 61  KYFVKKDLNRFDMYSLYALVASSEAICHANLNLDEIDADRFGVIVASGIGGIQEIEEQVI 120
            KYFVKKD NRFD YSLYALVA+ EA+ HANL+++ ++ DRFGVIVASGIGGI+EIE+QV+
Sbjct: 61  KYFVKKDTNRFDNYSLYALVAAQEAVNHANLDVEALNRDRFGVIVASGIGGIKEIEDQVL 120

Query: 121 RLHEKGPKRVKPMTLPKALPNMAAGNVAMRLGAHGVCKSINTACASSNDAIGDAFRNIKF 180
            RLHEKGPKRVKPMTLPKALPNMA+GNVAMR GA+GVCKSINTAC+SSNDAIGDAFR+IKF
Sbjct: 121 RLHEKGPKRVKPMTLPKALPNMASGNVAMRFGANGVCKSINTACSSSNDAIGDAFRSIKF 180

Query: 181 GIQDIMVVGGAEAAITKFAIAGFQSLTALSTTEDPSRASIPFDKDRNGFIMGEGSGNLVL 240
            G QD+M+VGG EA+IT FAIAGFQ+LTALSTTEDP+RASIPFDKDRNGF+MGEGSGNLVL
Sbjct: 181 GFQDVMLVGGTEASITPFAIAGFQALTALSTTEDPTRASIPFDKDRNGFVMGEGSGNLVL 240

Query: 241 ESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEGLGATKAIQLALVEANIKPEEVNYV 300
            ESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEG GA KAI+LAL EA I PE+V YV
Sbjct: 241 ESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEGQGAIKAIKLALEEAEISPEQVAYV 300

Query: 301 NAHGTSTPANEKGESQAIVAALGTDVPVSSTKSFTGHLLGAAGAVEAIATIEAIRHSYVP 360
            NAHGTSTPANEKGES AIVA LG +VPVSSTKSFTGHLLGAAGAVEAI TIEA+RH++VP
Sbjct: 301 NAHGTSTPANEKGESGAIVAVLGKEVPVSSTKSFTGHLLGAAGAVEAIVTIKAMRHNFVP 360

Query: 361 MTAGTTELSEDITANVIPGQGQDADIRYAISNTFGFGGHNAVLAFKRWED 410
            MTAGT+E+S+ I ANV++GQG + +I YAISNTFGFGGHNAVLAFKRWE+
Sbjct: 361 MTAGTSEVSDYIEANVVYGQGLEKEIPYAISNTFGFGGHNAVLAFKRWEN 410
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3851> which encodes the amino acid sequence <SEQ ID 3852>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.0890(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 346/410 (84%), Positives = 377/410 (91%)

Query: 1   MTLQRVVVTGYGVTSPIGNTPEEFWNSLKEGNVGIGPITKFDSSDPMVKNAAEIHDFPFD 60
           MT +RVVVTGYG+TSPIG+  PE FWN+LK G +GIGPITKFD++D+  VKNAAEI DFPFD
Sbjct: 1   MTFKRVVVTGYGLTSPIGHDPETFWNNLKAGQIGIGPITKFDTDYAVKNAAEIQDFPFD 60

Query: 61  KYFVKKDLNRPDMYSLYALYASSEAIQHANLNLDEIDADRVGVIVASGIGGIQEIEEQVI 120
           KYFVKKDLNRPD YSLYALYA+ EAI HA+LN++ +D+DRVGVIVASGIGGI EIEEQVI
Sbjct: 61  KYFVKKDLNRPDRYSLYALYAAKSAINHADLNIEMVDSDRFGVIVASGIGGIARIEEQVI 120

Query: 121 RLHEKGPKRVKPMTLPKALPNMAAGNVAMRLGAHGVCKSINTACASSNDAIGDAFRNIKF 180
           RLHEKGPKRVKPMTLPKALPNMAAGNVAM L A GVCKSINTACASSNDAIGDAFR IKF
Sbjct: 121 RLHEKGPKRVKPMTLPKALPNMAAGNVAMSLKAQGVCKSINTACASSNDAIGDAFRAIKF 180

Query: 181 GIQDIMVVGGAEAAITKFAIAGFQSLTALSTTEDPSRASIPFDKDRNGFIMGEGSGMLVL 240
           G QD+M+VGG+EAAITKFAIAGFQSLTALSTTEDPSR+SIPFDKDRNGFIMGEGSGMLVL
Sbjct: 181 GTQDVMIVGGSEAAITKFAIAGFQSLTALSTTEDPSRSSIPFDKDRNGFIMGEGSGMLVL 240

Query: 241 ESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEGLGNTKAIQLALVEANIKPEEVNYV 300
           ESLEHA++RGATILAE+VGYGNTCDAYHMTSP+PEGLGA KAI LAL EA I+   +NYV
Sbjct: 241 ESLEHAQERGATILAEIVGYGNTCDAYHMTSPNPKGLGARKAIHLALQEAGIEASAINYV 300

Query: 301 NAHGTSTPANEKGESQAIVAALGTDVPVSSTKSFTGHLLGAAGAVEALATIEAIRHSYVP 360
           NAHGTSTPANEKGESQAIVA LG DVPVSSTKSFTGHLLGAAGA+EALATIEA+RH+YVP
Sbjct: 301 NAHGTSTPANEKGESQAIVALGKDVPVSSTKSFTGHLLGAAGAIEALATIEAMRHNYVP 360

Query: 361 MTAGTTELSEDITANVIFGQGQDADIRYALSNTFGFGGHNAVLAFKRWED 410
           MTAGT  LSEDI ANVIFG+G++  I YALSNTFGFGGHNAVLAFK WE+
Sbjct: 361 MTAGTQALSEDIEANVIFGEGKSTAINYALSNTFGFGGHNAVLAFKCWEE 410
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1677

A DNA sequence (GBSx1781) was identified in *S.agalactiae* <SEQ ID 5201> which encodes the amino acid sequence <SEQ ID 5202>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3052(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9553> which encodes amino acid sequence <SEQ ID 9554> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98277 GB:AF197933 biotin carboxyl carrier protein
          [Streptococcus pneumoniae]
 Identities = 103/169 (60%), Positives = 127/169 (74%), Gaps = 11/169 (6%)

Query: 19  LDIQEIKDLMTQFDESSLREFSFKTSDGELSFSKNEGKADLVFTMSPMSHQPEATPTIAT 78
           +++ +IKDLMTQFD+SSLREFS+K    EL FSKNE +   VP ++ +  Q   P +AT
Sbjct: 1   MNLNDIKDLMTQFDQSSLREFSYKNGTDELQFSKNEARP--VPEVAT---QVAPAPVLAT 55

Query: 79  PVSNEAGEQTKQATEVVSEIP---ESTVTVAEGDVVESPLVGVAYLASGPDKPNFVSVGD 135
           P  +    + A  V  E+P   E++V    EG++VESPLVGV YLA+GPDKP FV+VGD
Sbjct: 56  P--SPVAPTSAPAETVAEVPAPAEASVAT-EGNLVESPLVGVVYLAAGPDKPAFVTVGD 112
```

```
Query: 136 SVKKGQTLMIIEAMKVMNEVPAPHDGVVTEILVANEEVIEPGKGLVRIK 184
           SVKKGQTL+IIEAMKVMNE+PAP DGVVTEILV+NEE++EPGKGLVRIK
Sbjct: 113 SVKKGQTLVIIEAMKVMNEIPAPKDGVVTEILVSNEEMVEPGKGLVRIK 161
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5203> which encodes the amino acid sequence <SEQ ID 5204>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3132(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 107/171 (62%), Positives = 126/171 (73%), Gaps = 10/171 (5%)

Query: 19  LDIQEIKDLMTQFDESSLREFSFKTSDGELSFSKNEGKAPLVPTMSPMSHQPEATPT--- 75
           L+IQEIKDLM QFD SSLREF FKT++GEL FSKNE              + S+Q  A P
Sbjct: 1   LNIQEIKDLNAQFDTSSLREFLFKTNEGELIFSKNEQHLN-----ASTSNQSHAVPVPQV 55

Query: 76  --IATPVSMEAGEQTKQATEVVSEIPESTVTVARGDVVESPLVGVAYLASGPDKPNFVSV 133
             +  P ++KA          V E P++   VAEGD+VESPLVGVAYLA+ PDKP FV+V
Sbjct: 56  QLVPNPTASKASSPASVKDVPVEEQPQASSFVAEGDIVESPLVGVAYLAASPDKPPFVAV 115

Query: 134 GDSVKKGQTLMIIEAMKVMNEVPAPHDGVVTEILVANEEVIEPGKGLVRIK 184
           GD+VKKGQTL+IIEAMKVMNEVPAP DGV+TEILV+NE+VIEPG+GLVRIK
Sbjct: 116 GDTVKKGQTLVIIEAMKVMNEVPAPCDGVITEILVSNEDVIEPGQGLVRIK 166
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1678

A DNA sequence (GBSx1782) was identified in *S.agalactiae* <SEQ ID 5205> which encodes the amino acid sequence <SEQ ID 5206>. This protein is predicted to be beta-hydroxyacyl-ACP dehydratase (fabZ). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2267(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98278 GB:AF197933 beta-hydroxyacyl-ACP dehydratase
          [Streptococcus pneumoniae]
 Identities = 130/140 (92%), Positives = 135/140 (95%)

Query: 1   MIDIKEIREALPHRYPMLLVDRVLEVSEDEIVAIKNVSINEPFFNGHFPEYPVMPGVLIM 60
           MIDI+ I+EALPHRYPMLLVDRVLEVSED IVAIKNV+INEPFFNGHFP+YPVMPGV+IM
Sbjct: 1   MIDIQGIKEALPHRYPMLLVDRVLEVSEDTIVAIKNVTINEPFFNGHFPQYPVMPGVVIM 60

Query: 61  EALAQTAGVLELSKEENKGKLVFYAGMDKVKFKKQVVPGDQLVMTAKFVKRRGTIAVVEA 120
           EALAQTAGVLELSK ENKGKLVFYAGMDKVKFKKQVVPGDQLVMTA FVKRRGTIAVVEA
Sbjct: 61  EALAQTAGVLELSKPENKGKLVFYAGMDKVKFKKQVVPGDQLVMTATFVKRRGTIAVVEA 120

Query: 121 IAEVDGKLAASGTLTFAIGN 140
            AEVDGKLAASGTLTFAIGN
```

```
Sbjct: 121 KAEVDGKLAASGTLTFAIGN 140
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5207> which encodes the amino acid sequence <SEQ ID 5208>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1882(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 127/139 (91%), Positives = 133/139 (95%)

Query: 1   MIDIKEIREALPHRYPMLLVDRVLEVSEDEIVAIKNVSINEPFFNGHFPEYPVMPGVLIM 60
           M+DI+EI+ ALPHRYPMLLVDRVLEVS+D IVAIKNV+INEPFFNGHFP YPVMPGVLIM
Sbjct: 1   MMDIREIQAALPHRYPMLLVDRVLEVSDDHIVAIKNVTINEPFFNGHFPHYPVMPGVLIM 60

Query: 61  EALAQTAGVLELSKEENKGKLVFYAGMDKVKPKKQVVPGDQLVMTAKFVKRRGTIAVVEA 120
           EALAQTAGVLELSKEENKGKLVFYAGMDKVKPKKQVVPGDQLVMTA F+KRRGTIAVVEA
Sbjct: 61  EALAQTAGVLELSKEENKGKLVFYAGMDKVKPKKQVVPGDQLVMTATFIKRRGTIAVVEA 120

Query: 121 IAEVDGKLAASGTLTFAIG 139
            AEVDGKLAASGTLTFA G
Sbjct: 121 RAEVDGKLAASGTLTFACG 139
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1679

A DNA sequence (GBSx1783) was identified in *S.agalactiae* <SEQ ID 5209> which encodes the amino acid sequence <SEQ ID 5210>. This protein is predicted to be acetyl-coenzyme A carboxylase, biotin carboxylase (accC). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1203(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98279 GB:AF197933 acetyl-CoA carboxylase biotin carboxylase
        subunit [Streptococcus pneumoniae]
 Identities = 361/451 (80%), Positives = 405/451 (89%)

Query: 1   MFKKILIANRGEIAVRIIRAAREMGISTVAIYSEADKESLHTLLADEAICVGPAKSAESY 60
           MF+KILIANRGEIAVRIIRAARE+GI+TVA+YS ADKE+LHT+LADEA+C+GP K+ ESY
Sbjct: 1   MFRKILIANRGEIAVRIIRAARELGIATVAVYSTADKRALHTLLADEAVCIGPGKATESY 60

Query: 61  LNVNAILSAAIVTGAEAVHPGFGFLSENSKFATMCEENNLKFIGPSGEVMDKMGDKINAR 120
           L+NA+LSAA++T AEA+HPGFGFLSENSKFATMCEE+ +KFIGPSG VMD MGDKINAR
Sbjct: 61  LHINAVLSAAVLTEAFAIHPGFGFLSENSKFATMCSEVGIKFIGPSGHVMDMMGDKINAR 120

Query: 121 TEMIKADVPVIPGSDGQVTSVKEAVSIAKEIGYPLMLKASAGGGGKGIRKVKSADELKPA 180
            +MIKA VPVIPGSDG+V + EEA+ +AE+IGYP+MLKASAGGGGKGIRKV+  D+L  A
Sbjct: 121 ACMIKAGVPVIPGSDGEVHNSKEALIVAEKIGYPVMLKASAGGGGKGIRKVEKPDDLVSA 180
```

```
Query: 181 FESASQEALAAFGNGAMYIEKVIYPARHIEVQILGDSFGKIVHLGERDCSLQRNNQKVLE 240
           FE+AS EA A +GNGAMYIE+VIYPARHIEVQILGD  G ++HLGERDCSLQRNNQKVLE
Sbjct: 181 FETASSEAKANYGNGAMYIERVIYPARHIEVQILGDEHGHVIHLGERDCSLQRNNQKVLE 240

Query: 241 ESPSVAIGNTLRQQIGEAAVRAAEAVSYENAGTIEFLLDENSGQFYFMEMNTRVQVEHPV 300
           ESPS+AIG TLR +IG AAVRAAE V YENAGTIEFLLDE S  FYFMEMNTRVQVEHPV
Sbjct: 241 ESPSIAIGKTLRHEIGAAAVRAABFVGYENAGTIEFLLDEASSNFYFMEMNTRVQVEHPV 300

Query: 301 TEFVTGVDIVKEQIRIAAGIPLSVSQNDIKLTGHAIECRINAENPQFNFAPCPGTINGLH 360
           TEFV+GVDIVKEQI IAAG PLSV Q DI L GHAIECRINAENP FNFAP PG I  L+
Sbjct: 301 TEFVSGVDIVKEQICIAAGQPLSVKQEDITLRGHAIECRINAENPAFNFAPSPGKITNLY 360

Query: 361 LPAGGMGLRVDSAVYTGYTIPPYYDSMIAKVIVHGENRPDALMKMQRALYELEIDGIVTN 420
           LP+GG+GLRVDSAVY GYTIPPYYDSMIAK+IVHGENRPDALMKMQRALYELEI+G+ TN
Sbjct: 361 LPSGGVGLRVDSAVYPGYTIPPYYDSMIAKIIVHGENRPDALMKMQRALYELEIEGVQTN 420

Query: 421 TEFQMDLISDKKVLAGDYDTSFLMEDFLPRY 451
           +FQ+DLISD+ V+AGDYDTSFLME FLP+Y
Sbjct: 421 ADFQLDLISDPNVIAGDYDTSFLMETFLPKY 451
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5211> which encodes the amino acid sequence <SEQ ID 5212>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1784(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 369/451 (81%), Positives = 421/451 (92%)

Query: 1   MFKKILIANRGEIAVRIIRAAREMGISTVAIYSEADKESLHTTILADEAICVGPAKSAESY 60
           MFKKILIANRGEIAVRIIRAARE+GISTVA+YSEADKE+LHTTILADEAIC+GPA+S ESY
Sbjct: 17  MFKKILIANRGEIAVRIIRAARELGISTVAVYSEADKEALHTTILADEAICIGPARSKESY 76

Query: 61  LNVNAILSAAIVTGAEAVHPGFGFLSENSKFATMCEENNLKFIGPSGEVMDKMGDKINAR 120
           LN+N++LSAAIVTGA+A+HPGFGFLSENSKFATMCEENN+KFIGPS  VMDKMGDKINAR
Sbjct: 77  LNMNSVLSAAIVTGAQAIHPGFGFLSENSKFATMCEENNIKFIGPSAASVMDKMGDKINAR 136

Query: 121 TEMIKADVPVIPGSDGQVTSVKEAVSIAKEIGYPLMLKASAGGGGKGIRKVKSADELKEA 180
           +EMIKA VPVIPGSDG+V + +EA++IA +IGYP+MLKASAGGGGKGIRKV++  +L+ A
Sbjct: 137 SEMIKAGVPVIPGSDGEVYNAQEALAIANKIGYPVMLKASAGGGGKGIRKVETEADLEAA 196

Query: 181 FESASQEALAAFGNGAMYIEKVIYPARHIEVQILGDSFGKIVHLGERDCSLQRNNQKVLE 240
           F +ASQEAL AFGNGAMY+EKVIYPARHIEVQILGD++G I+HLGERDCSLQRNNQKVLE
Sbjct: 197 FNAASQEALGAFGNGAMYLEKVIYPARHIEVQILGDAYGNIIHLGERDCSLQRNNQKVLE 256

Query: 241 ESPSVAIGNTLRQQIGEAAVRAAEAVSYENAGTIEFLLDENSKKFYFMEMNTRIQVEHPV 300
           ESPS+AIGNTLR ++G+AAVRAAEAV+YENAGTIEFLLDENSKKFYFMEMNTRIQVEHPV
Sbjct: 257 ESPSIAIGNTLRHEMGQAAVRAAEAVAYENAGTIEFLLDENSKKFYFMEMNTRIQVEHPV 316

Query: 301 TEFVTGVDIVKEQIRIAAGIPLSVSQNDIKLTGHAIECRINAENPQFNFAPCPGTINGLH 360
           TEFVTGVDIVKEQI+IAAG PL+++Q DI +TGHAIECRINAEN  FNFAP PG I  L+
Sbjct: 317 TEFVTGVDIVKEQIKIAAGQPLAINQEDITITGHAIECRINAENTAPNFAPSPGKITDLY 376

Query: 361 LPAGGMGLRVDSAVYTGYTIPPYYDSMIAKVIVHGENRPDALMKMQRALYELEIDGIVTN 420
           +P+GG+GLRVDSAVY GY IPPYYDSMIAK+IVHG NRPDALMKMQRAL ELEI+GI+TN
Sbjct: 377 MPSGGVGLRVDSAVYNGYAIPPYYDSMIAKIIVHGSNRPDALMKMQRALVELEIEGIITN 436

Query: 421 TEFQMDLISDKKVLAGDYDTSFLMEDFLPRY 451
           T+FQ+DLISDK+V+AGDYDTSFLME FLP Y
Sbjct: 437 TDFQLDLISDKRVIAGDYDTSFLMETFLPHY 467
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1680

A DNA sequence (GBSx1784) was identified in *S.agalactiae* <SEQ ID 5213> which encodes the amino acid sequence <SEQ ID 5214>. This protein is predicted to be acetyl-CoA carboxylase beta subunit (accD). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3571(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98280 GB:AF197933 acetyl-CoA carboxylase beta subunit
          [Streptococcus pneumoniae]
 Identities = 221/285 (77%), Positives = 248/285 (86%), Gaps = 1/285 (0%)

Query: 1   MALFSKKDKYIRISPNKALGSSDKRSLPEVPDELFAKCPSCKHMIYQKDLGLAKICPACS 60
           MALFSKKDKYIRI+PN++++    PEVPDELF++CP CKH IYQKDLG  +ICP CS
Sbjct: 1   MALFSKKDKYIRINPNRSVREKPQAK-PEVPDELFSQCPGCKHTIYQKDLGSERICPHCS 59

Query: 61  YNFRISAQERLLLTVDEDSFEELFTGIETKDPLNFPNYREKLAATRQKTNLDEAVVTGLA 120
           Y FRISAQERL LT+D  +F+ELFTGIE+KDPL+FP Y++KLA+ R+KT L EAVVTG A
Sbjct: 60  YTFRISAQERLALTIDMGTFKELFTGIESKDPLHFPGYQKKLASMREKTGLHEAVVTGTA 119

Query: 121 KIKGQTTALAIMDSHFIMASMGIVVGEKLTRLFELATEKKLPIVIFTASGGARMQEGIMS 180
            IKGQT AL IMDS+FIMASMGIVVGEK+TRLFE AT +KLP+V+FTASGGARMQEGIMS
Sbjct: 120 LIKGQTVALGIMDSNFIMASMGIVVGEKITRLFEYATVEKLPVVLFTASGGARMQEGIMS 179

Query: 181 LMQMAKVSAAVKRHSNQGLFYLTILTDPTTGGVTASFAMEGDIILAEPQAIVGFAGRRVI 240
           LMQMAK+SAAVKRHSN GLFYLTILTDPTTGGVTASFAMEGDIILAEPQ+LVGFAGRRVI
Sbjct: 180 LMQMAKISAAVKRHSNAGLFYLTILTDPTTGGVTASFAMEGDIILAEPQSLVGFAGRRVI 239

Query: 241 ETTVREDLPEGFQKAEFLLEHGFVDAIINRTELRDCLAQLIAFHG 285
           E TVRE LPE FQKAEFLLEHGFVDAI+ R +L D IA L+  HG
Sbjct: 240 ENTVRESLPEDFQKAEFLLEHGFVDAIVKRRDLDPDTIASLVRLHG 284
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5215> which encodes the amino acid sequence <SEQ ID 5216>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4092(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 232/285 (81%), Positives = 253/285 (88%)

Query: 1   MALFSKKDKYIRISPNKALGSSDKRSLPEVPDELFAKCPSCKHMIYQKDLGLAKICPACS 60
           MALF KKDKYIRI+PN +L S   ++PEVPDELFAKCP+CKHMIY+KDLGLAKICP CS
Sbjct: 1   MALFRKKDKYIRITPNNSLKGSVSHNVPEVPDELFAKCPACKHMIYKKDLGLAKICPTCS 60

Query: 61  YNFRISAQERLLLTVDEDSFEELFTGIETKDPLNFPNYREKLAATRQKTNLDEAVVTGLA 120
           YNFRISAQERL LTVDE SF+ELFT IETKDPL FP Y+EKL   ++ T L EAV+TG A
```

```
Sbjct: 61  YNFRISAQERLTLTVDEGSFQELFTSIETKDPLRFPGYQRKLQKAKETTGLHEAVLTGKA 120

Query: 121 KIKGQTTALAIMDSHFIMASMGTVVGEKLTRLFELATEKKLPVIVFTASGGARMQEGIMS 180
            +K Q  ALAIMDSHFIMASMGTVVGEK+TRLFELA E+ LP+VIFTASGGARMQEGIMS
Sbjct: 121 MVKEQKIALAIMDSHFIMASMGTVVGEKITRLFELAIENLPVVIFTASGGARMQEGIMS 180

Query: 181 LMQMAKVSAAVKRHSNQGLFYLTILTDPTTGGVTASFAMEGDIILAEPQALVGFAGRRVI 240
            LMQMAKVSAAVKRHSN GLFYLTILTDPTTGGVTASFAMEGDIILAEPQ+LVGFAGRRVI
Sbjct: 181 LMQMAKVSAAVKRHSNAGLFYLTILTDPTTGGVTASFAMEGDIILAEPQSLVGFAGRRVI 240

Query: 241 ETTVREDLPEGFQKAEFLLEHGFVDAIINRTELRDCIAQLIAFHG 285
            ETTVRE+LP+ FQKAEFL +HGFVDAI+ RTELRD IA L+AFHG
Sbjct: 241 ETTVRENLPDDFQKAEFLQDHGFVDAIVKRTELRDKIAHLVAFHG 285
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1681

A DNA sequence (GBSx1785) was identified in *S.agalactiae* <SEQ ID 5217> which encodes the amino acid sequence <SEQ ID 5218>. This protein is predicted to be acetyl-CoA carboxylase alpha subunit (accA). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22   Transmembrane  149 - 165 ( 149 - 165)

----- Final Results -----
               bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9555> which encodes amino acid sequence <SEQ ID 9556> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98281 GB:AF197933 acetyl-CoA carboxylase alpha subunit
            [Streptococcus pneumoniae]
 Identities = 186/254 (73%), Positives = 222/254 (87%)

Query: 13  DVTRILKDARDQGRLTALDYAELIPDNFMELHGDRQFADDKSIIGGLGYLAGRPVTIVGI 72
            ++ +I+++AR+Q RLT LD+A  IPD F+LHGDR F DD +++GG+G+L  + VT+VGI
Sbjct: 2   NIAKIVREARERQGRLTTLDFATGIPDEFIQLHGDRSFRDDGAVVGGIGWLGDQAVTVVGI 61

Query: 73  QKGKNLQDNLDRHFGQPHPEGYRKALRLMKQAEKFGRPVITFINTAGAYPGVGAEERGQG 132
            QKGK+LQDNL R+FGQPHPEGYRKALRLMKQAEKFGRPV+TFINTAGAYPGVGAEERGQG
Sbjct: 62  QKGKSLQDNLKRNFGQPHPEGYRKALRLMKQAEKFGRPVVTFINTAGAYPGVGAEERGQG 121

Query: 133 EAIARNLLEMSDLKVPIIAIIIGEGGSGGALALAVADKVWMLEHTVYSILSPEGFASILW 192
            EAIARNL+EMSDLKVPIIAIIIGEGGSGGALALAVAD+VWMLE+++Y+ILSPEGFASILW
Sbjct: 122 EAIARNLMEMSDLKVPIIAIIIGEGGSGGALALAVADRVWMLENSIYAILSPEGFASILW 181

Query: 193 KDGTRTTEAAQLMKMTAGELYHMEVVDKVIPEHGYFSSEIVDMIKTSLISELEVLSQLSL 252
            KDGTR  EAA+LMK+T+ EL  M+VVDKVI E G  S E++  +K  L +EL  LSQ  L
Sbjct: 182 KDGTRAMEAAELMKITSHELLEMDVVDKVISEIGLSSKELIKSVKKELQTELARLSQKPL 241

Query: 253 EDLLEQRYQRFKKY 266
            E+LLE+RYQRFKKY
Sbjct: 242 EELLEERYQRFKKY 255
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5219> which encodes the amino acid sequenee <SEQ ID 5220>. Analysis of this protein sequence reveals the following:

```
     Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22   Transmembrane  139 - 155 ( 139 - 155)

----- Final Results -----
                bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF98281 GB:AF197933 acetyl-CoA carboxylase alpha subunit
             [Streptococcus pneumoniae]
 Identities = 185/254 (74%), Positives = 225/254 (88%)

Query: 3   DVSRILKEARDQGRLVTLDYANLIPDDFMELHGDRHFSDDGAIVGGLAYLAGQPVTVIGI 62
            ++++I++EAR+Q RLTTLD+A  IPD+F++LHGDR F DDGA+VGG+ +L  Q VTV+GI
Sbjct: 2   NIAKIVREAREQSRLVTLDFATGIPDEFIQLHGDRSFRDDGAVVGGIGWLGDQAVTVVGI 61

Query: 63  QKGKNLQDNLARNFGQPNPEGYRKALRLMKQAEKFGRPVVTFINTAGAYPGVGAEERGQG 122
            QKGK+LQDNL RNFGQP+PEGYRKALRLMKQAEKFGRPVVTFINTAGAYPGVGAEERGQG
Sbjct: 62  QKGKSLQDNLKRNFGQPHPEGYRKALRLMKQAEKFGRPVVTFINTAGAYPGVGAEERGQG 121

Query: 123 EAIAKNLMEMSDLKVPIIAIIIGEGGSGGALALAVADQVWMLENTMYAVLSPEGFASILW 182
            EAIA+NLMEMSDLKVPIIAIIIGEGGSGGALALAVAD+VWMLEN++YA+LSPEGFASILW
Sbjct: 122 EAIARNLMEMSDLKVPIIAIIIGEGGSGGALALAVADRVWMLENSIYALLSPEGFASILW 181

Query: 183 KDGSRATEAAELMKITAGELYKMGIVDRIIPEHGYFSSEIVDIIKANLIEQITSLQAKPL 242
            KDG+RA EAAELMKIT+ EL +M +VD++I E G  S E++  +K  L  ++  L  KPL
Sbjct: 182 KDGTRAMEAAELMKITSHELLEMDVVDKVISEIGLSSKELIKSVKKELQTELARLSQKPL 241

Query: 243 DQLLDERYQRPRKY 256
            ++LL+ERYQRPRKY
Sbjct: 242 ERLLEERYQRPRKY 255
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 204/254 (80%), Positives = 236/254 (92%)

Query: 13  DVTRILKDARDQGRLTALDYAELIPDNFMELHGDRQFADDKSIIGGLGYLAGRPVTIVGI 72
            DV+RILK+ARDQGRLT LDYA LIPD+FMELHGDR F+DD +I+GGL YLAG+PVT++GI
Sbjct: 3   DVSRILKEARDQGRLTTLDYANLIPDDFMELHGDRHFSDDGAIVGGLAYLAGQPVTVIGI 62

Query: 73  QKGKNLQDNLDRHFGQPHPEGYRKALRLMKQAEKFGRPVITFINTAGAYPGVGAEERGQG 132
            QKGKNLQDNL R+FGQP+PEGYRKALRLMKQAEKFGRPV+TFINTAGAYPGVGAEERGQG
Sbjct: 63  QKGKNLQDNLARNFGQPNPEGYRKALRLMKQAEKFGRPVVTFINTAGAYPGVGAEERGQG 122

Query: 133 EAIARNLLEMSDLKVPIIAIIIGEGGSGGALALAVADKVWMLEHTVYSILSPEGFASILW 192
            EAIA+NL+EMSDLKVPIIAIIIGEGGSGGALALAVAD+VWMLE+T+Y++LSPEGFASILW
Sbjct: 123 EAIAKNLMEMSDLKVPIIAIIIGEGGSGGALALAVADQVWMLENTMYAVLSPEGFASILW 182

Query: 193 KDGTRTTEAAQLMKMTAGELYHMEVVDKVIPEHGYFSSEIVDMIKTSLISELEVLSQLSL 252
            KDG+R TEAA+LMK+TAGELY M +VD++IPEHGYFSSEIVD+IK +LI ++  L    L
Sbjct: 183 KDGSRATEAAELMKITAGELYKMGIVDRIIPEHGYFSSEIVDIIKANLIEQITSLQAKPL 242

Query: 253 EDLLEQRYQRPRKY 266
            + LL++RYQRPRKY
Sbjct: 243 DQLLDERYQRPRKY 256
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1682

A DNA sequence (GBSx1786) was identified in *S.agalactiae* <SEQ ID 5221> which encodes the amino acid sequence <SEQ ID 5222>. This protein is predicted to be sakacin A production response regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3304(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9557> which encodes amino acid sequence <SEQ ID 9558> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA88824 GB:AB016077 sakacin A production response regulator
          [Streptococcus mutans]
 Identities = 76/142 (53%), Positives = 99/142 (69%)

Query: 36  MQTPKAKGQLARNSFTELSRALEQRMDGFKMQRVSWNANQAQVGRPHFWVYYRKDTDQLD 95
           M   K  GQ AR  FTE+++ L   ++  F+M RVSWNANQAQV RPHFW YY++  D  D
Sbjct: 1   MIALKTLGQSARAEFTEIAKVLALKVSPFEMMRVSWNANQAQVVRPHFWCYYKQFEDNQD 60

Query: 96  DVAVALRVYGVKDSPGVSLKVSFVERQKSDKTLEKQARVLSIPIASPLYFMVQRQGETHR 155
           DV +A+R+YG   +FG+S+EVSF+ER+KS  TL KQ +VL IPIA PLY+  Q + E+HR
Sbjct: 61  DVGLAIRLYGNSANFGISVEVSPIERKKSKATLAKQHKVLDIPIASPLYYFAQEKSESHR 120

Query: 156 EEGNEENRQRLMQEIKSGKVRK 177
            G E  RQ L Q++  G+VRK
Sbjct: 121 VSGTEAYRQMLRQKVADGQVRK 142
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1683

A DNA sequence (GBSx1787) was identified in *S.agalactiae* <SEQ ID 5223> which encodes the amino acid sequence <SEQ ID 5224>. This protein is predicted to be seryl-tRNA synthetase (serS). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1866(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11789 GB:Z99104 seryl-tRNA synthetase [Bacillus subtilis]
 Identities = 262/425 (61%), Positives = 322/425 (75%), Gaps = 1/425 (0%)

Query: 1   MLDLKRIRTDPDVVAKKLATRGVDQETLITLKELDIKRRELLIKAEEAKAQRNVASAAIA 60
           MLD K +R +F  +  KL  +G D        + LD +RREL+ K EE K +RN  S  +A
Sbjct: 1   MLDTKMLRANPQEIKAKLVHKGEDLTFDKPFEALDDRRRELIGKVEELKGKRNEVSQQVA 60
```

```
Query: 61  QAKRNKENADEQIAAMQTLSADIKAIDARLADVDANLQSMVTVLPNTPADDVPLGADEDE 120
             KR K++AD  I  M+ +  +IK +D EL  V+A L +++  +PN P + VP+G  ED+
Sbjct: 61  VLKREKKDADHIIKEMREVGEEIKKLDERLRTVEARLDTILLSIPNIPHESVPVGETEDD 120

Query: 121 NVEVRRWGTPREFDFETKAHWDLGESLGILDWERGAKVTGSRFLFYKGLGARLERAIYSF 180
            NVEVR+WG     F +E K HWD+ + LGILD+ER AKVTGSRF+FYKGLGARLERA+Y+F
Sbjct: 121 NVEVRKWGKPSFAYEPKPHWDIADELGILDFERAAKVTGSRFVFYKGLGARLERALYNF 180

Query: 181 MLDEHAKR-GYTEVIPPYMVNHDSMPGTGQYPKFKEDTFELADSPFVLIPTAEVPLTNYY 239
            MLD H  E  YTEVIPPYMVN  SM GTGQ PKF+ED F++ +  + LIPTAEVP+TN +
Sbjct: 181 MLDLHVDEYNYTEVIPPYMVNRASMTGTGQLPKFEEDAFKIRDEDYFLIPTAEVPTTNMH 240

Query: 240 RDEIIDGKELPIYPTAMSPSFRSEAGSAGRDTRGLIRLHQPHKVEMVKFAKPREESYQELE 299
            RDEI+ G  LPI + A S  FRSEAGSAGRDTRGLIR HQF+KVE+VKF KPR+SY+ELE
Sbjct: 241 RDEILSGDSLPINYAAFSACFRSEAGSAGRDTRGLIRQHQFNKVELVKFVKPEDSYEELE 300

Query: 300 KMTANAENILQKLNLPYRVITLCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQ 359
            K+T  AE +LQ L LPYRV+++CTGD+GF+AAK YD+EVWIP+Q+TYREISSCSN E FQ
Sbjct: 301 KLTNQAERVLQLLELPYRVMSMCTGDLGFTAAKKYDIRVWIPSQDTYREISSCSNFEAFQ 360

Query: 360 ARRAQIRYRDSVDGKVRLLHTLNGSGLAVGRTVAAILENYQNEDGSVTIPEVLRPYMGNI 419
            ARRA IR+R E  GK   +HTLNGSGLAVGRTVAAILENYQ EDGSV IP+VLRPYMGN
Sbjct: 361 ARRANIRFRREAKGKPEHVHTLNGSGLAVGRTVAAILENYQQEDGSVVIPKVLRPYMGNR 420

Query: 420 DIIKP 424
            +++KP
Sbjct: 421 EVMKP 425
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5225> which encodes the amino acid sequence <SEQ ID 5226>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2453(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 357/424 (84%), Positives = 386/424 (90%)

Query: 1   MLDLKRIRTDFDVVAKKLATRGVDQETLTTLKELDIKRRELLIKAEEAKAQRNVASAAIA 60
            MLDLKRITDFD VA KL  RGV ++TLT LKELD KRR LL++EE KA+RN+ASAAIA
Sbjct: 1   MLDLKRITDFDTVAAKLKNRGVSEDTLTHLKELDSKRRALLVQSEELKAERNIASAAIA 60

Query: 61  QAKRNKENADEQIAAMQTLSADIKAIDARLADVDANLQSMVTVLPNTPADDVPLGADEDE 120
            QAKR KE+A +QIA MQ +SADIK ID +L  +D  +  ++TVLPNTP D VP+GADE++
Sbjct: 61  QAKRQKEDATQQIADMQKVSADIKTIDNQLVAIDQQVTDIITTVLPNTPHDSVPVGADEED 120

Query: 121 NVEVRRWGTPREFDFETKAHWDLGESLGILDWERGAKVTGSRFLFYKNLGARLERALTYSF 180
            NVE+RRWGTPR+FDFE KAHWDLGE L ILDWERGAKVTG+RFLFYK LGARLERA+Y+F
Sbjct: 121 NVEIRRWGTPRDFDFEVKAHWDLGEDLDTILDWERGAKVTGARFLFYKNLGARLERALNYF 180

Query: 181 MLDEHAKEGYTEVIPPYMVNHDSMPGTGQYPKFKEDTFELADSPFVLIPTAEVPLTNYYR 240
            MLDEH KEGY E+I PYMVNHDSMPGTGQYPKFKEDTFELAD+ FVLIPTAEVPLTNYYR
Sbjct: 181 MLDEHIKEGYQEIITPYMVNHDSMPGTGQYPKFKEDTFELADTNFVLIPTAEVPLTNYYR 240

Query: 241 DEIIDGKELPIYPTAMSPSFRSEAGSAGRDTRGLIRLHQFHKVEMVKFAKPREESYQELEK 300
             EI+DGKELPIYPTAMSPSFRSEAGSAGRDTRGLIRLHQFHKVEMVKFAKPREESYQELEK
Sbjct: 241 GEILDGKELPIYPTAMSPSFRSEAGSAGRDTRGLIRLHQFHKVEMVKFAKPREESYQELEK 300

Query: 301 MTANAENILQKLNLPYRVITLCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQA 360
            MTANAENILQKL LPYRVI+LCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQA
Sbjct: 301 MTANAENILQKLGLPYRVISLCTGDMGFSAAKTYDLEVWIPAQNTYREISSCSNTEDFQA 360
```

```
Query: 361 RRAQIRYRDEVDGKVRLLHTLNGSGLAVGRTVAAILENYQNEDGSVTIPEVLRPYMGNID 420
           RRAQIRYRDE DGKV+LLHTLNGSGLAVGRTVAAILENYQNEDGSVTIPEVLRPYMG
Sbjct: 361 RRAQIRYRDEADGKVKLLHTLNGSGLAVGRTVAAILENYQNEDGSVTIPEVLRPYMGGET 420

Query: 421 IIKP 424
           +I P
Sbjct: 421 VISP 424
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1684

A DNA sequence (GBSx1788) was identified in *S.agalactiae* <SEQ ID 5227> which encodes the amino acid sequence <SEQ ID 5228>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.36   Transmembrane  313 - 329 ( 306 - 332)
   INTEGRAL    Likelihood = -9.24   Transmembrane  159 - 175 ( 155 - 179)
   INTEGRAL    Likelihood = -4.19   Transmembrane   20 -  36 (  16 -  37)
   INTEGRAL    Likelihood = -3.29   Transmembrane  271 - 287 ( 271 - 287)
   INTEGRAL    Likelihood = -2.97   Transmembrane  210 - 226 ( 209 - 227)
   INTEGRAL    Likelihood = -2.87   Transmembrane  242 - 258 ( 241 - 258)
   INTEGRAL    Likelihood = -2.13   Transmembrane   52 -  68 (  50 -  68)

----- Final Results -----
               bacterial membrane --- Certainty=0.5543(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9559> which encodes amino acid sequence <SEQ ID 9560> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA07406 GB:AJ006986 transmembrane protein [Streptococcus pneumoniae]
 Identities = 72/330 (21%), Positives = 143/330 (42%), Gaps = 32/330 (9%)

Query: 14  RHYGLDLLRIISMFMIVITHVLGKGGLRSSVEGHADSYFIVTWIIQVLVYGAVNCYALIS 73
           R+  LDLL+++    +V+ H    GG + +   +  +Y            + ++   VN Y L+
Sbjct: 5   RNINLDLLKVLACVGVVLLHTT-MGGFKETGANNFLTYLYYLGTYSIPLFFMVNGYLLL- 62

Query: 74  GYVGINSRYRYSKLLSIWAQVFFYTFTITALFATTGHE------VTLLNWRDAFFPIVSG 127
           G   I   Y  K+  +   V  +TF I  LF      E      +  L  +  FP
Sbjct: 63  GKREITYSYILQKIKWLLITVSSWTF-IVWLFKDFTENLIKKIIGSLIQKGYFP----- 116

Query: 128 QYWYITAYFGLLVFMPVINNGLNALTDKQLKQLVLLMFI--IFSILPAVLNNRVPEFSLS 185
           Q+W+  A   + + +P++   LN+     L  L LLM I  IF +   +L   +  + +
Sbjct: 117 QFWFFGALILIYLCLPILRQFLNS-KRSYLYSLSLLMTIGLIFELSNILLQMPIQTYVIQ 175

Query: 186 KGPEMTWLLILVLIGAYLKRIDL----NLFKTSYLLIYYLLSLVATYAMKFSVGDIW--- 238
               TW    Y++G Y+ +  +    + FK    ++  LL L++   + F    I+
Sbjct: 176 TFRLWTW-FFYYLLGGYIAQFTIESISRFKNWMKIVSILLLLISPILFFIAKTIYHNL 234

Query: 239 ---YWYVSPTLTLGAVSLFILFARASIKPSGFLKKIIVVLAPSTLGVVLCHLIPLIVKYF 295
              Y+Y +  + +  + +F+     ++  +  ++ IV L+  T+GV++  +H  I+K +
Sbjct: 235 FAEYFYDTLFVKVSTLGIFLTILMLTLNEN--RRESIVESLSNQTMGVPI--IHTYIMKVW 290

Query: 296 VRDFARTFVYESIYLYPFLILGAGILIYLL 325
           +    FV   +  F +  + I++ +L
Sbjct: 291 EKVLGFNFVGAYLLFALFTLSVSFIIVGML 320
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1685

A DNA sequence (GBSx1789) was identified in *S.agalactiae* <SEQ ID 5229> which encodes the amino acid sequence <SEQ ID 5230>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2752(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9561> which encodes amino acid sequence <SEQ ID 9562> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD46488 GB:AF130465 unknown [Streptococcus salivarius]
 Identities = 88/112 (78%), Positives = 96/112 (85%)

Query: 1   MAQSLNKTVSPQTTGVSYLGMGNKVGKFLWGDQALEFYNDKNVNDYIQIPWTSINQIGAN 60
            MAQSLNKTVS  TTGVSY+ +G KVGKFL+GD ALEFY D NV  YIQIPWTSI QIGAN
Sbjct: 1   MAQSLNKTVSLHTTGVSYMAIGGKVGKFLIGDVALEFYFDVNVEQYIQIPWTSITQIGAN 60

Query: 61  VSRKKISRHFEVFTDQGKFLFASKDSGTILKHARRHIGDDKVVKLPTLIQTI 112
            VS K+ISRHFEV TD+ KFLFASKDSG ILK AR H+G++KVVKLPTLIQTI
Sbjct: 61  VSGKRISRHFEVLTDKSKFLFASKDSGKILKIARKHLGNEKVVKLPTLIQTI 112
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5231> which encodes the amino acid sequence <SEQ ID 5232>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3301(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 87/116 (75%), Positives = 101/116 (87%)

Query: 1   MAQSLNKTVEPQTTGVSYLGMGNKVGKFLWGDQALEFYNDKNVNDYIQIPWTSINQIGAN 60
            MAQSLN +VE++T  VSYLGMG KVG  L+GD+ALEFYNDKNVNDYIQIPWT+IN IGAN
Sbjct: 1   MAQSLNTSVEYKTKAVSYLGMGGKVGHILLGDKALEFYNDKNVNDYIQIPWTAINHIGAN 60

Query: 61  VSRKKISRHFEVFTDQGKFLFASKDSGTILKHARRHIGDDKVVKLPTLIQTILKTF 116
            VSRKK+SRHFE+FTDQGKFLFAS DSG ILK  R+HIG++KV+ LPTL+QT +  F
Sbjct: 61  VSRKKVSRHFEIFTDQGKFLFASGDSGKILKITRQHIGNEKVITLPTLMQTFINKF 116
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1686

A DNA sequence (GBSx1790) was identified in *S.agalactiae* <SEQ ID 5233> which encodes the amino acid sequence <SEQ ID 5234>. This protein is predicted to be mannose-specific phosphotransferase system component IID (manZ). Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.92   Transmembrane  281 - 297 ( 279 - 302)
   INTEGRAL    Likelihood = -4.88   Transmembrane  187 - 203 ( 185 - 205)
   INTEGRAL    Likelihood = -4.35   Transmembrane  260 - 276 ( 257 - 277)
   INTEGRAL    Likelihood = -1.01   Transmembrane  129 - 145 ( 129 - 145)

----- Final Results -----
                bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD46487 GB:AF130465 mannose-specific phosphotransferase system
             component IID [Streptococcus salivarius]
 Identities = 247/303 (81%), Positives = 276/303 (90%)

Query: 1   MTEQIKLSKSDRQKVWWRSQFLQGSWNYERMQNWGWAYALIPALKKLYTTKEDRAAALER 60
            M E+I+LS++DR+KVWWRSQFLQGSWNYERMQN+GWAY+LIPA+KKLYT KED+AAALR
Sbjct: 1   MAEKIQLSQADRKKVWWRSQFLQGSWNYERMQNLGWAYSLIPAIKKLYTNKEDQAAALKR 60

Query: 61  HMEFFNTHPYVAAPIIGVTLALEEEKASGTPVEDKAIQGVKIGMMGPLAGIGDPVFWFTV 120
            H+EFFNTHPYVAAPI+GVTLALEEEKA+GT +ED AIQGVKIGMMGPLAGIGDPVFWFTV
Sbjct: 61  HLEFFNTHPYVAAPIMGVTLALEEEKANGTDIEDAAIQGVKIGMMGPLAGIGDPVFWFTV 120

Query: 121 RPILGALGASLASAGNILGPIIPFVGWNLIRMSPLWYTQELGYKSGKEITKDMSGGILQD 180
            RPILGALGASLA AGNI GP+IPF+GWNLIRM+PLWYTQELGYK+G EITKDMSGGILKD
Sbjct: 121 RPILGALGASLAQAGNIAGPLIPFIGWNLIRMAPLWYTQELGYKAGSEITKDMSGGILKD 180

Query: 181 ITKGASILGMPILALVLVKRWVAINFTVDLPKKTLSEGAYINFPKDMVSGQQLWDILGQVQ 240
            ITKGASILGMPILALVL+RWV+I FTV+LP K LS+GAYI +PK +VSG QL  ILGQV
Sbjct: 181 ITKGASILGMPILALVLERWVSIVFTVNLPGKVLSKGAYIEWPKGNVSGDQLKTTLGQVN 240

Query: 241 SGLSLDKMQPQTLQSQLDSLIPGLAGLLLTFPCMWLLKKKVSPITIIIGLPIVGILARLA 300
              LS DK+Q  TLQ QLDSLIPGL GLLLTF CMWLLKKKVSPITIIIGLP+VGI+A
Sbjct: 241 DKLSFDKIQVDTTLQKQLDSLIPGLMGLLLTFACMWLLKKKVSPITIIIGLFVVGIVASFF 300

Query: 301 GVM 303
            G+M
Sbjct: 301 GIM 303
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5235> which encodes the amino acid sequence <SEQ ID 5236>. Analysis of this protein sequence reveals the following:

```
      Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.39   Transmembrane  284 - 300 ( 279 - 302)
   INTEGRAL    Likelihood = -4.88   Transmembrane  261 - 277 ( 257 - 278)
   INTEGRAL    Likelihood = -4.51   Transmembrane  181 - 197 ( 180 - 198)

----- Final Results -----
                bacterial membrane --- Certainty=0.4354(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD46487 GB:AF130465 mannose-specific phosphotransferase system
             component IID [Streptococcus salivarius]
```

```
 Identities = 239/303 (78%), Positives = 268/303 (87%)

Query: 1   MTEQIKLTKSDRQRVWWRSQFLQGSWNYERMQNMGWAYALIPALKKLYTSPEDRAAALER 60
            M E+I+L++DR++VWWRSQFLQGSWNYERMQN+GWAY+LIPA+KKLYT+ ED+AAAL+R
Sbjct: 1   MAEKIQLSQADRKKVWWRSQFLQGSWNYERMQNLGWAYSLIPAIKKLYTNKEDQAAALKR 60

Query: 61  HMEPFNTHPYVAAPIIGVTLALEEERANGTPIDDKAIQGVKIGMMGPLAGIGDPVFWFTI 120
            H+EPFNTHPYVAAPI+GVTLALEEE+ANGT I+D AIQGVKIGMMGPLAGIGDPVFWFT+
Sbjct: 61  HLEPFNTHPYVAAPIMGVTLALEEKKANGTDIEDAAIQGVKIGMMGPLAGIGDPVFWFTV 120

Query: 121 RPILGALGASLASTGNIVGPLLFFPGNNLIRMAPLWYTQEFGYKAGSEITKDMSGGILQD 180
            RPILGALGASLA  GNI GPL+FP GNNLIRMAPLWYTQE GYKAGSEITKDMSGGIL+D
Sbjct: 121 RPILGALGASLAQAGNIAGPLIFFIGNNLIRMAPLWYTQRLGYKAGSRITKDMSGGILKD 180

Query: 181 ITKGASILGMPILAVLVQRWVSINPTIDLPGKQLSDGAYVVPFDGAVKGAELKTILANAI 240
            ITKGASILGMPILAVLV+RWVS+ PT++LPGK LS GAY+ +P G V G +LKTIL
Sbjct: 181 ITKGASILGMPILAVLVERWVSIVPTVNLPGKVLSKGAYIEWFKGNVSGDQLKTILGQVN 240

Query: 241 GGMSLDKVQAQTLQGQLDSLIPGLAGLLLTFLCMWLLKKKVSPIAIIIGLFAPGILAHLA 300
              +S DK+Q  TLQ QLDSLIPGL GLLLTF CMWLLKKKVSPI IIIGLF  GI+A
Sbjct: 241 DKLSFDKIQVDTLQKQLDSLIPGLMGLLLTFACMWLLKKKVSPITIIIGLFVVGIVASFF 300

Query: 301 GIM 303
            GIM
Sbjct: 301 GIM 303
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 255/303 (84%), Positives = 277/303 (91%)

Query: 1   MTEQIKLSKSDRQKVWWRSQFLQGSWNYERMQNMGWAYALIPALKKLYTTKEDRAAALER 60
            MTEQIKL+KSDRQ+VWWRSQFLQGSWNYERMQNMGWAYALIPALKKLYT+ EDRAAALER
Sbjct: 1   MTEQIKLTKSDRQRVWWRSQFLQGSWNYERMQNMGWAYALIPALKKLYTSPEDRAAALER 60

Query: 61  HMEPFNTHPYVAAPIIGVTLALEEEKASGTPVEDKAIQGVKIGMMGPLAGIGDPVFWFTV 120
            HMEPFNTHPYVAAPIIGVTLALEEE+A+GTP++DKAIQGVKIGMMGPLAGIGDPVFWFT+
Sbjct: 61  HMEPFNTHPYVAAPIIGVTLALEEERANGTPIDDKAIQGVKIGMMGPLAGIGDPVFWFTI 120

Query: 121 RPILGALGASLASAGNILGPIIFFVGNNLIRMSPLWYTQELGYKSGKEITKDMSGGILQD 180
            RPILGALGASLAS GNI+GP++FP GNNLIRM+PLWYTQE GYK+G EITKDMSGGILQD
Sbjct: 121 RPILGALGASLASTGNIVGPLLFFPGNNLIRMAPLWYTQERGYKAGSEITKDMSGGILQD 180

Query: 181 ITKGASILGMPILAVLVRWVAINPTVDLPKKTLSEGAYINFPKDHVSGQQLHDILGQVQ 240
            ITKGASILGMPILAVLV+RWV+INPT+DLP K LS+GAY+ FP   V G +L  IL
Sbjct: 181 ITKGASILGMPILAVLVQRWVSINPTIDLPGKQLSDGAYVVPFDGAVKGAELKTILANAI 240

Query: 241 SGLSLDKMQPQTLQGQLDSLIPGLAGLLLTFPCMWLLKKKVSPITIIIGLFIVGILARLA 300
             G+SLDK+Q QTLQGQLDSLIPGLAGLLLTF CMWLLKKKVSPI IIIGLF  GILA LA
Sbjct: 241 GGMSLDKVQAQTLQGQLDSLIPGLAGLLLTFLCMWLLKKKVSPIAIIIGLFAPGILAHLA 300

Query: 301 GVM 303
            G+M
Sbjct: 301 GIM 303
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1687

A DNA sequence (GBSx1791) was identified in *S.agalactiae* <SEQ ID 5237> which encodes the amino acid sequence <SEQ ID 5238>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
bacterial cytoplasm --- Certainty=0.2580(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1688

A DNA sequence (GBSx1792) was identified in *S.agalactiae* <SEQ ID 5239> which encodes the amino acid sequence <SEQ ID 5240>. This protein is predicted to be mannose-specific phosphotransferase system component IIC (manY). Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.95   Transmembrane  142 - 158 ( 137 - 165)
   INTEGRAL    Likelihood = -2.60   Transmembrane   65 -  81 (  61 -  81)
   INTEGRAL    Likelihood = -1.97   Transmembrane  103 - 119 ( 103 - 122)

----- Final Results -----
                bacterial membrane --- Certainty=0.3378(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9301> which encodes amino acid sequence <SEQ ID 9302> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD46486 GB:AP130465 mannose-specific phosphotransferase system
            component IIC [Streptococcus salivarius]
 Identities = 134/186 (72%), Positives = 154/186 (82%), Gaps = 1/186 (0%)

Query: 1   MVKSGDPTQKGINFAFSTAVPLAIAGLFLTMIVRTISTALVHAGDKAASEGNFAAIERFH 60
           +VK G+PT +GI  A +TA+PLA+AGLFLTM+VRT S ALVHA DKAA  GN A +ER H
Sbjct: 86  LVKGGNPTTEGIGVATATAIPLAVAGLFLTMLVRTASVALVHAADKAAESGNIAGVERAH 145

Query: 61  FIALLLQGLRIAFPAALLLAIPSSSVQSILEAMPDWLNGGMQVGGAMVVAVGYAMVINNM 120
           ++ALLLQGLRIA PAALLAIP+ SVQ  L  MP WLN GM VG3 MVVAVGYAMVINNM
Sbjct: 146 YLALLLQGLRIAVPAALLAIPAESVQHALGLMPSWLNHGMVVGGSMVVAVGYAMVINNM 205

Query: 121 ATREVWPFFALGFALAALNQLTLIAMGTIGVAIALIYISLSKMGSSK-GTSNAGSNDPIG 179
           ATREVWPFFA+GFA AA++QLTLIA+G IGVAIA IY++LSK GG   G +++GS DPIG
Sbjct: 206 ATREVWPFFAIGFAFAAISQLTLALGAIGVAIAPIYLNLSKQGGNGGGTSSGSGDPIG 265

Query: 180 DILEDY 185
           DILEDY
Sbjct: 266 DILEDY 271
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5241> which encodes the amino acid sequence <SEQ ID 5242>. Analysis of this protein sequence reveals the following:

```
     Possible site: 36
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.30   Transmembrane    4 -  20 (   1 -  28)
   INTEGRAL    Likelihood = -7.64   Transmembrane  226 - 242 ( 212 - 247)
   INTEGRAL    Likelihood = -4.14   Transmembrane  102 - 118 ( 101 - 123)
   INTEGRAL    Likelihood = -3.77   Transmembrane   71 -  87 (  69 -  87)
   INTEGRAL    Likelihood = -3.40   Transmembrane  150 - 166 ( 146 - 167)
```

```
   INTEGRAL    Likelihood = -2.13   Transmembrane  186 - 202 ( 186 - 202)
   INTEGRAL    Likelihood = -0.37   Transmembrane   37 -  53 (  37 -  53)

  ----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD46486 GB:AF130465 mannose-specific phosphotransferase system
        component IIC [Streptococcus salivarius]
 Identities = 211/271 (77%), Positives = 237/271 (86%), Gaps = 2/271 (0%)

Query: 1   MSDISIISAILVVIIAFFAGLEGILDQFQMHQPLVACTLIGLVTGHLEAGVILGGTLQML 60
           MSD+SIISAILVV++AF AGLEGILDQFQ HQPLVACTLIG  TG+L AG++LGG+LQM+
Sbjct: 1   MSDMSIISAILVVVVAFLAGLEGILDQFQPHQPLVACTLIGAATGNLTAGIMLGGSLQMI 60

Query: 61  ALGWANIGAAVAPDAALASVAAAIIMVKSGDFTQKGITFAYSTAIPLAVAGLFLTMIVRT 120
           AL WANIGAAVAPDAALASVAAAII+VK G+FT +GI  A +TAIPLAVAGLFLTM+VRT
Sbjct: 61  ALAWANIGAAVAPDAALASVAAAIILVKGGNFTTEGIGVATATAIPLAVAGLFLTMLVRT 120

Query: 121 LSTAIVHAGDKAAESGNPAGIERFHFIALLLQGLRIAVPAALLVAVPTSAVQSVLNAMPN 180
            S AIVHA DKAA  GN AG+ER H++ALLLQGLRIAVPAALL+A+P  +VQ  L  MP+
Sbjct: 121 ASVALVHAADKAAESGNIAGVERARYLALLLQGLRIAVPAALLAIPAESVQRALGLMPS 180

Query: 181 WLNEGMQIGGAMVVAVGYAMVINMMATREVWPFFALGFALAAISQLTLIAMGVIGVAIAF 240
           WLN GM +GG MVVAVGYAMVINMMATREVWPFFA+GFA AAISQLTLIA+G IGVAIAF
Sbjct: 181 WLNHGMVVGGGMVVAVGYAMVINMMATREVWPFFAIGFAFAAISQLTLIALGAIGVAIAF 240

Query: 241 IYLNLSKKGG--NGGNAAGSADPIGDILEDY 269
           IYLNLSK+GG   GG ++GS DPIGDILEDY
Sbjct: 241 IYLNLSKQGGGNGGGTSSGSGDPIGDILEDY 271
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 155/185 (83%), Positives = 173/185 (92%), Gaps = 1/185 (0%)

Query: 1   MVKSGDFTQKGINFAPSTAVPLAIAGLFLTMIVRTISTALVHAGDKAASEGNFAAIERFH 60
           MVKSGDFTQKGI FA+STA+PLA+AGLFLTMIVRT+STALVHAGDKAA+EGNFA IERFH
Sbjct: 86  MVKSGDFTQKGITFAYSTAIPLAVAGLFLTMIVRTLSTALVHAGDKAAAEGNTAGIERFH 145

Query: 61  FIALLLQGLRIAFPAALLLAIPSSSVQSILEAMPDWLNGGMQVGGAMVVAVGYAMVINMM 120
           FIALLLQGLRIA PAALL+A+P+S+VQS+L AMP+WLN GMQ+GGAMVVAVGYAMVINMM
Sbjct: 146 FIALLLQGLRIAVPAALLVAVPTSAVQSVLNAMPNWLNEGMQIGGAMVVAVGYAMVINMM 205

Query: 121 ATREVWPFFALGFALAAINQLTLIAMGTIGVAIALIYISLSKMGGSKGTSNAGSNDPIGD 180
           ATREVWPFFALGFALAA++QLTLIAMG IGVAIA IY++LSK GG+ G + AGS DPIGD
Sbjct: 206 ATREVWPFFALGFALAAISQLTLIAMGVIGVAIAFIYLNLSKKGGNGGNA-AGSADPIGD 264

Query: 181 ILEDY 185
           ILEDY
Sbjct: 265 ILEDY 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1689

A DNA sequence (GBSx1793) was identified in *S.agalactiae* <SEQ ID 5243> which encodes the amino acid sequence <SEQ ID 5244>. Analysis of this protein sequence reveals the following:

```
  Possible site: 37
  >>> Seems to have no N-terminal signal sequence

  ----- Final Results -----
```

```
bacterial cytoplasm --- Certainty=0.3171(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1690

A DNA sequence (GBSx1794) was identified in *S.agalactiae* <SEQ ID 5245> which encodes the amino acid sequence <SEQ ID 5246>. This protein is predicted to be pseudouridine synthase (rluC). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2717(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06566 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 124/281 (44%), Positives = 171/281 (60%), Gaps = 8/281 (2%)

Query: 16  LLKSHDVSRGLLAKIKYRGGKIPVNGEEQNAIPLLEIGDVVTIDIPDE-PSHETL-EPVP 73
           L +   VS+  LA IK++GG I +NGEE      + + D VT+++P E PS   + EPVP
Sbjct: 24  LREGKHVSKRSLAAIKFKGGTIILNGEEVTVRETVHVNDQVTLELPHEYPSPSMIAEPVP 83

Query: 74  HDLDIIYEDDHFLIINKPPGFASIPSSIH-SNTIANFIGIYYVSNNYANQQVHIVTRLDR 132
             D+IYE+DH+L++NKP G  +IPS  H   T+NAN + +Y+     A     H V RLD+
Sbjct: 84  --PDVIYENDHYLVVNKPAGVPTIPSRDHPQSTLANGLLNYFQRQKMA-ATFHAVNRLDK 140

Query: 133 DTSGLMLFAKHGYAHARLDKQLQAKAIEKRYYALVSGSGDLADSGDIIAPIARDVDSIIT 192
           DTSGL++ AKH  AH +L KQ +   I++ Y A+V G  +  + G I APIAR  +S+IT
Sbjct: 141 DTSGLLIVAKHQLAHDQLSKQQRQGNIKRTYMAIVQGEIEQQE-GTITAPIARKEESLIT 199

Query: 193 RRVHESGKYAHTSYQVVAKYGDVRLVDIKLHTGRTHQIRVHFAHIGFPLLGDDLYGGRMD 252
           R V E G+ A T ++V+ R      +V ++L TGRTHQIRVHF+++G+PL GDDLYGG
Sbjct: 200 REVREDGQLAITHFKVIDRLNQGTIVQVQLETGRTHQIRVHFSYLGYPLVGDDLYGGRRK 259

Query: 253 LGINRQALHCHSLSFVDPFMGKINKQTLDLTDDFDSVIMKL 293
            GI RQALH   L+ + PF          T  L  D   +I  L
Sbjct: 260 -GIERQALHSTELTIHCPFTEVEQTFTEGLPFDMKELIRHL 299
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5247> which encodes the amino acid sequence <SEQ ID 5248>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2786(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 223/294 (75%), Positives = 251/294 (84%), Gaps = 1/294 (0%)
```

```
Query: 1   MKFEYVAKERCKVKTLLKSHDVSRGLLAKIKYRGGKIFVNGEEQNAIFLLEIGDVVTIDI 60
           M+FE+VA +R KVKTLLKS+DVS+GLLAKIKY+GG I VNG EQNAI+LL++GDVVTIDI
Sbjct: 1   MRFEFVADKRIKVKTLLKSYDVSKGLLAKIKYKGGNILVNGIEQNAIYLLQVGDVVTIDI 60

Query: 61  PDEPSHSETLEPVPHDLDIIYEDDHFLILNKPPGFASIPSSIHSNTIANFIKHYYVSNNYA 120
           P+E   E LS +P DLDI++KDDHFL++NKP GFASIPS+IHSNTIANFIK YYV N+Y
Sbjct: 61  PNESPFEKLRAIPPDLDIVHSDDHFLVINKPIGFASIPSAIHSNTIANFIKAYYVDNHYL 120

Query: 121 NQQVHIVTRLDRDTSGLMLFAKHGYAHARLDKQLQAKAIEKRYYALVSGSGDLADSGDII 180
           +QQVHIVTRLDRDTSGLMLFAKHGYAHARLDKQLQ ++IEKRY+ALVSG+G L D GDII
Sbjct: 121 DQQVHIVTRLDRDTSGLMLFAKHGYAHARLDKQLQTRSIEKRYFALVSGNGMLPDEGDII 180

Query: 181 APIARDVDSIITRRVHESGKYAHTSYQVVARYGD-VRLVDIKLHTGRTHQIRVHFAHIGF 239
           API R  DSIITR V   GKYA TSY+VVARY + V LVDIKLETGRTHQIRVHFAHIGF
Sbjct: 181 APIGRSKDSIITRAVDEWGKYAKTSYKVVARYSENVHLVDIKLHTGRTHQIRVHFAHIGF 240

Query: 240 PLLGDDLYGGRMDLGINRQALHCHSLSFYDPFMGKINKQTLDLTDDFDSVIMEL 293
           PLLGDDLYGGR+DLGI RQALHCH  L+F DPF        + LTDDFDSVI+ L
Sbjct: 241 PLLGDDLYGGRLDLGITRQALHCHYLNFKDPFTESDCSYATHLTDDFDSVIIGL 294
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1691

A DNA sequence (GBSx1795) was identified in *S.agalactiae* <SEQ ID 5249> which encodes the amino acid sequence <SEQ ID 5250>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1521(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9845> which encodes amino acid sequence <SEQ ID 9846> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13018 GB:Z99110 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 120/267 (44%), Positives = 174/267 (64%), Gaps = 3/267 (1%)

Query: 13  RVAIIANGKYQSKRVASKLFAAFKHDPDFYLSKKDPDIVISIGGDGMLLSAFHMYEKQLD 72
           + A+ + G   S  + SK+  A+  D D  L  + +P+IVIS+GGDG LL AFH Y  +LD
Sbjct: 2   KFAVSSKGDQVSDTLKSKI-QAYLLDPMKLDENSPSIVISVGGDGTLLYAFHRYSDRLD 60

Query: 73  KVRFVGVHTGHLGFYTDYRDPEVDTLINNLKNDKGEQISYPILKVTITL-EDGKVIRARA 131
           K  FVGVHTGHLGFY D+   E++ L+  +          + YP+L+V +T  E+ R  R A
Sbjct: 61  KTAFVGVHTGHLGFYADWVPHEIEKLVLAIAKTPYHTVEYPLLSEVIVTVHENRERERYLA 120

Query: 132 LNESTCKRIEKTMVADVVINQVVPERPRGDGILVSTPTGSTAYNKSLGGAVLHPTIEALQ 191
           LNE TIK IE ++VADV I   +PE PRGDG+ +STP+GSTAYNK+LGGA++HP+I A+Q
Sbjct: 121 LNRCTLKSIRGSLVADVEIKGQLPETPRGDGLCLSYPSGSTAYNKALGGAIIHPSIRAIQ 180

Query: 192 LTEISSLNNRVVRTLGSSVIIPKVDAIEIVPERVGVYTTSIDNKTVHYKNVTKIEYSIDE 251
           L E++S+NNRV+RT+GS +++P     I P+    + ++ID+ T+ +K+V  I   +
Sbjct: 181 LAEMASINNRVFRTVGSPLILPSHDCMIKPRNEVDFQVTIDHLTLLHKDVKSIRCQVAS 240

Query: 252 KSINFVSTPSHTSFWERVNDAFIGEPE 278
           + + F         FW+RV D+FIG+ E
Sbjct: 241 EKVRFARFRPF-PFWKRVQDSPIGKGE 266
```

A related sequence was also identified in GAS <SEQ ID 9137> which encodes the amino acid sequence <SEQ ID 9138>. Analysis of this protein sequence reveals the following:

```
      Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2190(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 155-157
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 232/276 (84%), Positives = 257/276 (93%)

Query: 1   MMTQMNFTDRATRVAIIANGKYQSKRVASKLFAAFKHDPDFYLSKKDPDIVISIGGDGML 60
           +MTQMN+T +  RVAIIANGKYQSKRVASKLF+ FK DPDFYLSKK+PDIVISIGGDGML
Sbjct: 1   VMTQMNYTGKVKRVAIIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGML 60

Query: 61  LSAFHMYEKQLDKVRFVGVHTGHLGFYTDYRDFEVDTLINNLKNDKGEQISYPILKVTIT 120
           LSAFHMYEK+LDKVRFVG+HTGHLGFYTDYRDFEVD LI+NL+ DKGEQISYPILKV IT
Sbjct: 61  LSAFHMYEKELDKVRFVGIHTGHLGFYTDYRDFEVDKLIDNLRKDKGEQISYPILKVAIT 120

Query: 121 LEDGRVIRARALNESTIKRIEKTMVADVVINQVVFERFRGDGILVSTPTGSTAYNKSLGG 180
           L+DGRV++ARALNE+T+KRIEKTMVADV+IN V FE FRGDGI VSTPTGSTAYNKSLGG
Sbjct: 121 LDDGRVVKARALNEATVKRIEKTMVADVIINEVKFESFRGDGISVSTPTGSTAYNKSLGG 180

Query: 181 AVLHPTIEALQLTEISSLNNRVYRTLGSSVIIPKKDAIEIVPKRVGVVTISIDNKTVHYK 240
           AVLHPTIEALQLTEISSLNNRV+RTLGSS+IIPKKD IE+VPKR+G+VTISIDNKT   K
Sbjct: 181 AVLHPTIEALQLTRISSLNNRVFRTLGSSIIIPKKDKIELVPKRLGIVTISIDNKTYQLK 240

Query: 241 NVTKIEYSIDEKSINFVSTPSHTSPWSRVNDAFIGE 276
           NVTK+EY ID++ I+FVS+PSHTSPWSRV DAFIGE
Sbjct: 241 NVTKVEYFIDDEKIHFVSSPSHTSPWSRVKDAFIGE 276
```

A related GBS gene <SEQ ID 8879> and protein <SEQ ID 8880> were also identified. Analysis of this protein sequence reveals an RGD motif at residues 159-161.

The protein has homology with the following sequences in the databases:

```
45.0/65.6% over 264aa
                                                          Bacillus subtilis

  EGAD|107338| hypothetical protein Insert characterized OMNI|NT01BS1363 BC541A protein-
related Insert characterized
  SP|O31612|YJEN_BACSU HYPOTHETICAL 30.0 KDA PROTEIN IN MECA-TRNA INTERGENIC REGION. Insert
characterized
  GP|2633515|emb|CAB13018.1||Z99110 similar to hypothetical proteins Insert characterized
  PIR|F69844|F69844 conserved hypothetical protein yjbN - Insert characterized

ORF02026(337 - 1134 of 1437)
EGAD|107338|BS1162(2 - 266 of 266) hypothetical protein (Bacillus subtilis) OMNI|NT01BS1363
BC541A protein-related SP|O31612|YJEN_BACSU HYPOTHETICAL 30.0 KDA PROTEIN IN MECA-TRNA
INTERGENIC REGION. GP|2633515|emb|CAB13018.1||Z99110 similar to hypothetical proteins
(Bacillus subtilis) PIR|F69844|F69844 conserved hypothetical protein yjbN - Bacillus
subtilis
%Match = 22.8
%Identity = 44.9  %Similarity = 65.5
Matches = 120  Mismatches = 89  Conservative Sub.s = 55

87        117       147       177       207       237       267       297
RKF*QKYKSELWL*IFGQPSNIH*ITSIRGTSLKKLNKDWRKQQKSL*NWMKKCVRFAKIFVKHSFYLIL*IEN*AMV*E

327       357       387       417       447       477       507       537
```

```
IVMTQMNFTDRATRVAIIANGKYQSKRVASKLFAAPKHDPDFYLSKKDPDIVISIGGDGMLLSAFHNYEKQLDKVRFVGV
                : |: : |   |  : ||: | :  | |  | : :|:|||:|||| || ||| |  :||| ||||
              MKPAVSSKGDQVSDTLKSKIQA-YLLDFDMELDENEPKIVISVGGDGTLLYAFHRYSDRLDKTAFVGV
                        10        20        30        40        50        60

567       597       627       657       684       714       744       774
HTGHLGFYTDYRDFEVDTLINNLKNDKGEQISYPILKVTTTL-EDGRVIRARALNESTIKRIEKTMVADVVINQVVFERF
|||||||| |:   |:: |:  :      : ||:|:| :|  |: |  |  |||| ||| || ::|||| |    :|| |
HTGHLGFVADWVPHEIKKLVLAIAKTPYHTVEYPLLSVIVTYHENKREKRYIALNKCTIKSIEGSLVADVEIKGQLFETF
          80        90       100       110       120       130       140

804       834       864       894       924       954       984       1014
RGDGILVSTPTGSTAYNKSLGGAVLHPTIRALQLTEISSLNNRVYRTLGSSVIIPKKDAIEIVPKRVGVYTISIDNKTVH
||||: :|||:|||||||:||||::||:| |:|| |::|:||||:||:|| :::|      | |:     : ::||: |:
RGDGLCLSTPSGSTAYNKALGGAIIHPSIRAIQLAEMASINNRVFRTVGSPLLLPSHHDCMIKPRNEVDFQVTIDHLTLL
          160       170       180       190       200       210       220

1044      1074      1104      1134      1164      1194      1224      1254
YKNVTKIEYSIDEKSINFVSTPSHTSPWERVNDAFIGKPRE*NLWT*QKKGAKLKHP*KVMMFQGGY*QRLSTEVVRFLL
:|:|  |   :  : : | :      ||:|| |:||| |
HKDVKSIRCQVASEKVRF-ARPRPFPPWKRVQDSFIGKGE
          240       250       260
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5251> which encodes the amino acid sequence <SEQ ID 5252>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2190(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS sequences follows:

```
 Score =  481 bits (1224), Expect = e-138
 Identities = 233/276 (84%), Positives = 257/276 (92%)

Query: 1   VMTQMNYTGKVKRVAIIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGML 60
           VMTQMN+T +  RVAIIANGKYQSKRVASKLF+ FK DPDFYLSKK+PDIVISIGGDGML
Sbjct: 1   VMTQMNPTDRATRVAIIANGKYQSKRVASKLFAAFKHDPDFYLSKKDPDIVISIGGDGML 60

Query: 61  LSAFHMYEKKLDKVRFVGIHTGHLGFYTDYRDFEVDKLIDNLRKDKGEQISYPILKVAIT 120
           LSAFHMYEK+LDKVRFVG+HTGHLGFYTDYRDFEVD LI+NL+ DKGEQISYPILKV IT
Sbjct: 61  LSAFHMYEKQLDKVRFVGVHTGHLGFYTDYRDFEVDTLINNLKNDKGEQISYPILKVTIT 120

Query: 121 LDDGRVVKARALNEATVKRIEKTMVADVIINHVKFESRGDGISVSTPTGSTAYNKSLGG 180
           L+DGRV++ARALNE+T+KRIEKTMVADV+IN V FS RGDGI VSTPTGSTAYNKSLGG
Sbjct: 121 LEDGRVIRARALNESTIKRIEKTMVADVVINQVVFERFRGDGILVSTPTGSTAYNKSLGG 180

Query: 181 AVLHPTIRALQLTEISSLNNRVFRTLGSSIIPKKDKIKLVPKRLGIYTTISIDNKTYQLK 240
           AVLHPTIRALQLTEISSLNNRV+RTLGSS+IIPKKD IE+VPKR+G+YTISIDNKT   K
Sbjct: 181 AVLHPTIEALQLTEISSLNNRVYRTLGSSVIIPKKDAIEIVPKRVGVYTISIDNKTVHYK 240

Query: 241 NVTKVEYPIDDEKIHFVSSPSHTSPWERVKDAFIGK 276
           NVTK+EY ID++ I+FVS+PSHTSPWERV DAFIGK
Sbjct: 241 NVTKIEYSIDEKSINFVSTPSHTSPWERVNDAFIGK 276
```

SEQ ID 8880 (GBS308) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 4; MW 34kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 3; MW 59kDa).

GBS308-GST was purified as shown in Figure 226, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1692

A DNA sequence (GBSx1796) was identified in *S.agalactiae* <SEQ ID 5253> which encodes the amino acid sequence <SEQ ID 5254>. This protein is predicted to be permease. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3653(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06568 GB:AP001516 GTP pyrophosphokinase [Bacillus halodurans]
 Identities = 115/208 (55%), Positives = 159/208 (76%), Gaps = 3/208 (1%)

Query: 4   DWETFLDPYIQTVGELKIKLRGIRKQPRKQNRHSPIEFVTGRVKSVESIQEKMVLRGISE 63
           +W+ FL PY Q V ELK+KL+GIR+Q+K ++H+PIEFVTGRVK + SI +K + + I
Sbjct: 3   NWDVFLTPYKQAVEELKVKLKGIREQYQKSSKHTPIEFVTGRVKPISSILDKAIRKNIPL 62

Query: 64  ENLAQDLQDIAGLRIMVQFVDDVDEVLALLRKRHDMTVVQERDYITHMKSSGYRSYHVVV 123
           + L + +QD+AGLRI+ QFV+D++ V+ L+R R D  +V+ERDY+   K SGYRSYH+V+
Sbjct: 63  DQLEEKMQDLAGLRIVTQFVEDIETVVQLIRSRSDFEIVERDYVEQKKDSGYRSYHLVL 122

Query: 124 EYPVDTIDGQKKVLAEIQIRTLAMNFWATIEHSLNYKYQGDFPEEIKQRLEKTAKIALEL 183
            YPV TI+G+K++L E+QIRTLAMNFWATIEHSLNYKY G+ P  IK RL++ A+ A  L
Sbjct: 123 RYPVQTIEGEKRILVELQIRTLAMNFWATIEHSLNYKYGGEIPLNIKTRLQRAAEAAFRL 182

Query: 184 DEEMRKIREDIREAQLLFDPLNRKLSDG 211
           DEEM +IR+++REAQ +   + RK   G
Sbjct: 183 DEEMSQIRDEVREAQQI---ITRKQEQG 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5255> which encodes the amino acid sequence <SEQ ID 5256>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4064(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 196/223 (87%), Positives = 213/223 (94%)

Query: 1   MSMDWETFLDPYIQTVGELKIKLRGIRKQPRKQNRHSPIEFVTGRVKSVRSIQEKMVLRG 60
           M++DWE FLDPYIQTVGELKIKLRGIRKQ+RKQNR+SPIEFVTGRVKS+ESI+EKM+LRG
Sbjct: 1   MTLDWEEFLDPYIQTVGELKIKLRGIRKQYRKQNRYSPIEFVTGRVKSIESIKEKMILRG 60

Query: 61  ISEENLAQDLQDIAGLRIMVQFVDDVDEVLALLRKRHDMTVVQERDYITHMKSSGYRSYH 120
           + EEN+AQD+QDIAGLRIMVQFVDDV+EVLALLR+R DMT+V ERDYI +MKSSGYRSYH
Sbjct: 61  VIEENIAQDIQDIAGLRIMVQFVDDVEEVLALLRQRQDMTIVYERDYIRNMKSSGYRSYH 120

Query: 121 VVVEYPVDTIDGQKKVLAEIQIRTLAMNFWATIEHSLNYKYQGDFPEEIKQRLEKTAKIA 180
           VVVEYPVDTI+GQKKVLAEIQIRTLAMNFWATIEHSLNYKY GDFPEEIK+RLE TAKIA
Sbjct: 121 VVVEYPVDTIEGQKKVLAEIQIRTLAMNFWATIEHSLNYKYGGDFPEEIKKRLEVTAKIA 180
```

```
Query: 181 LELDEEMRKIREDIREAQLLFDPLNRKLSDGVGNSDDTDEFYR 223
           LELDEEMRKIREDIREAQLLFDP+ R LSDGVGNSDDTDE YR
Sbjct: 181 LELDEEMRKIREDIREAQLLFDPVTRNLSDGVGNSDDTDELYR 223
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1693

A DNA sequence (GBSx1797) was identified in *S.agalactiae* <SEQ ID 5257> which encodes the amino acid sequence <SEQ ID 5258>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2266(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13015 GB:Z99110 yjbK [Bacillus subtilis]
 Identities = 63/184 (34%), Positives = 99/184 (53%), Gaps = 10/184 (5%)

Query: 4   LEIEYKTLLNKDEFNRLTSLFSHVQP--ITQTNYYFDTETFEMKAHRMSLRIRTLPNRAE 61
           +EIE+K +L K EF  + S      +     Q N+YFDT++F +K     LRIR     +
Sbjct: 5   IEIEFKNMLTKQEFKNIASALQLTEKDFTDQKNHYFDTDSFALKQKHAALRIRRKNGKYV 64

Query: 62  LYLKIPREVGNLEHNHDLT--LEEAKYIVKNGQPFEDTEIASLILEKGVDPTKLAVFGQL 119
           LYLK P +VG LE +  L+    + A + V  G  P    ++   L    +D    + FG L
Sbjct: 65  LYLKEPADVGLLETHQQLSEVSDLAGFSVPEG--FVKDQLHKL----QIDTDAIQYFGSL 118

Query: 120 TTTRREMETSIGLMALDSNIYADIKDYELELEVKQPKQGKRDPDQFLKENNININFKYAKSK 179
            T R E ET  GL+ LD + Y + +DYE+E E      +G++ F++ L++ +I  +  K+K
Sbjct: 119 ATNRAEKETEKGLIVLDHSRYLNKEDYEIEFEAADWHEQRQAFEKLLQQFSIPQRETYNK 178

Query: 180 VARF 183
           + RF
Sbjct: 179 ILRF 182
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5259> which encodes the amino acid sequence <SEQ ID 5260>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3470(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 114/188 (60%), Positives = 139/188 (73%), Gaps = 1/188 (0%)

Query: 1   MTHLEIEYKTLLNKDEFNRLTSLFSHVQPITQTNYYFDTETFEMKAHRMSLRIRTLPNRA 60
           MT+LEIEYKTLL K+E+NRL S   HV P+TQTNYY DT+ F++KA++MSLRIRT  N A
Sbjct: 1   MTNLEIEYKTLLTKNEYNRLLSQMKHVTPVTQTNYYIDTKAFDLKANKMSLRIRTFVNSA 60

Query: 61  ELTLKIPREVGNLEHNHDLTLEEAKYIVKNGQPFEDTEIASLILEKGVDPTKLAVFGQLT 120
           ELTLK+P +VGN E+N  L LE+AK ++K+G  PE T +  +I+ KG+ P+ L  FG LT
Sbjct: 61  ELTLKVPEKVGNREYNVFLPLEQAKDMIKHGNLPESTAL-DIIISKGIKPSALVTFGNLT 119
```

```
Query: 121 TTREEMETSIGLMALDSNIYADIKDYELELKVKQPKQGKRDFDQFLKENNINFKYAKSKV 180
           T RRE    IG +ALD N+YA+ KDYELELEV    QGK DFD FL E +I FKYAKSKV
Sbjct: 120 TVRETVIPIGKLALDYNLYANTKDYELELEVSDALQGKIDFDSFLSEYHITFKYAKSKV 179

Query: 181 ARFSATLK 188
           AR   TLK
Sbjct: 180 ARCINTLK 187
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1694

A DNA sequence (GBSx1798) was identified in *S.agalactiae* <SEQ ID 5261> which encodes the amino acid sequence <SEQ ID 5262>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1815(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1695

A DNA sequence (GBSx1799) was identified in *S.agalactiae* <SEQ ID 5263> which encodes the amino acid sequence <SEQ ID 5264>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0621(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1696

A DNA sequence (GBSx1800) was identified in *S.agalactiae* <SEQ ID 5265> which encodes the amino acid sequence <SEQ ID 5266>. This protein is predicted to be ribose-phosphate pyrophosphokinase (prsA). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
bacterial cytoplasm --- Certainty=0.3369(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11827 GB:Z99104 phosphoribosyl pyrophosphate synthetase
[Bacillus subtilis]
 Identities = 166/319 (52%), Positives = 231/319 (72%), Gaps = 4/319 (1%)

Query: 1   MAEQVADKQIKLFSLTANREIAEKISQASGIPLGKMSSRQFSDGEIMINIEETVRGDDIY 60
           M+ QY DK +K+FSL +N E+A++I+   G+ LGK S  +FSDGE+ INIEE++RG D Y
Sbjct: 1   MSNQYGDKNLKIFSLNSNPELAKEIADIVGVQLGKCSVTRFSDGEVQINIEESIRGCDCY 60

Query: 61  IIQSTSFPVNDNLWELLIMIDACKRASANTVNIVVPYFGYSRQDRIAASREPITAKLVAN 120
           IIQSTS PVN+++ ELLIM+DA KRASA T+NIV+PY+GY+RQDR A SREPITAKL AN
Sbjct: 61  IIQSTSDPVNKHIMELLIMVDALKRASAKTINIVIPTYGYARQDRKARSREPITAKLFAN 120

Query: 121 MLVKAGVDRVLTLDLHAVQVQGFFDIPVDNLFTVPLFAEHYNQLGLSGEDVVVVSPKNSG 180
           +L  AG  RV+ LDLHA Q+QGFFDIP+D+L  VP+  E++   G + ED+V+VSP + G
Sbjct: 121 LLETAGATRVIALDLHAPQIQGFFDIPIDHLMGVPILGEYFE--GKNLEDIVIVSPDHGG 178

Query: 181 IKRARSLAEYLDSPIAIIDYAQD-DSRREGYIIGEVEGKKAIIIDDIIATGKTFAEAAK 239
           + RAR LA+ L +PIAIID  +  +  E   I+G +EGK AI+IDDI++T  T  AA
Sbjct: 179 VTRARKLADRLKAPIAIIDKRRPRPNVAEVMNIVGNIEGKTAILIDDIIDTAGTITLAAN 238

Query: 240 ILERGGATEIYAVASHGLFAGGAADILESAPIREIIVTDSV-LSKERIPSNIKYLTASHL 298
             L   GA E+YA  +H + +G A  + +  ++ I+E++VT+S+ L +E+     K L+   L
Sbjct: 239 ALVENGAKEVYACCTHPVLSGPAVERINNSTIKELVVTNSIKLPEEKKIERFKQLSVGPL 298

Query: 299 IADAIIRIHERKPLSPLFS 317
           +A+AIIR+HE++ +S LFS
Sbjct: 299 LAEAIIRVHEQGSVSYLFS 317
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5267> which encodes the amino acid sequence <SEQ ID 5268>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1830(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 278/324 (85%), Positives = 305/324 (93%)

Query: 1   MAEQVADKQIKLFSLTANREIAEKISQASGIPLGKMSSRQFSDGEIMINIEETVRGDDIY 60
           M E+YADKQIKLFSLT+N  IAEKI++A+GIPLGKMSSRQFS+GEIMINIEETVRGDDIY
Sbjct: 1   MTERYADKQIKLFSLTSNLPIAEKIAKAAGIPLGKMSSRQFSNGEIMINIEETVRGDDIY 60

Query: 61  IIQSTSFPVNDNLWELLIMIDACKRASANTVNIVVPYFGYSRQDRIAASREPITAKLVAN 120
           IIQSTSFPVNDNLWELLIMIDACKRASANTVNIV+PYFGYSRQDR+A  REPITAKLVAN
Sbjct: 61  IIQSTSFPVNDNLWELLIMIDACKRASANTVNIVLPYFGYSRQDRKAPREPITAKLVAN 120

Query: 121 MLVKAGVDRVLTLDLHAVQVQGFFDIPVDNLFTVPLFAEHYNQLGLSGEDVVVVSPKNSG 180
           ML KAG+DRV+TLDLHAVQVQGFFDIPVDNLFTVPLFAE Y++LGLSG DVVVVSPKNSG
Sbjct: 121 MLTKAGIDRVVTLDLHAVQVQGFFDIPVDNLFTVPLFAEKYSKLGLSGSDVVVVSPKNSG 180

Query: 181 IKRARSLAEYLDSPIAIIDYAQDDSERESGYIIGEVRGKKAIIIDDIIATGKTFAEAAKI 240
           IKRARSLAEYLDSPIAIIDYAQDDSERE+GYIIG+V GKKAI+IDDILNTGKTFAEAAKI
Sbjct: 181 IKRARSLAEYLDSPIAIIDYAQDDSERESGYIIGDVSGKKAILIDDILNTGKTFAEAAKI 240

Query: 241 LERGGATEIYAVASHGLFAGGAADILESAPIREIIVTDSVLSKERIPSNIKYLTASHLIA 300
           LER GAT+ YAVASHGLFAGGAAD+LE+API+EIIVTDSV +K R+P N+ YL+AS LIA
```

```
Sbjct: 241 LERSGATDTYAVASHGLFAGGAADVLETAPIKEIIVTDSVKTKNRVPENVTYLSASDLIA 300

Query: 301 DAIIRIHERKPLSPLFSYRSDKKD 324
           +AIIRIHER+PLSPLFSY+   K+
Sbjct: 301 EAIIRIHERRPLSPLFSYQPKGKN 324
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1697

A DNA sequence (GBSx1801) was identified in *S.agalactiae* <SEQ ID 5269> which encodes the amino acid sequence <SEQ ID 5270>. This protein is predicted to be Fe-S cluster formation protein. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1981(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04979 GB:AP001511 Fe-S cluster formation protein [Bacillus halodurans]
 Identities = 174/373 (46%), Positives = 237/373 (62%), Gaps = 6/373 (1%)

Query: 3   IYLDNAATTALTPSVIEKMTKVMTSNYGNPSSIHTFGRQANQLLRECRQIIASYLNVNSR 62
           IYLD+AAT+ + P VI+ M       +GNPSSIH FGR+A Q + E R  IA  L  +
Sbjct: 4   IYLDHAATSPVHPEVIQAMLPYYEEQFGNPSSIHQFGRRARQGVDEARGTIARLLQADPS 63

Query: 63  EIIFTSGGTESNNTAIKGYALANQLKGKHIITSRIEHHSVLHTMTYLSERPGFDITYLKP 122
           E IFTSGGTE++N AI GYA  ++ KG HIITS++EHH+VLH    L E  GF++TY+
Sbjct: 64  EFIFTSGGTEADNLAIRGYAYQHRGKGNHIITSQVEHHAVLHACQEL-EHQGFEVTYVPV 122

Query: 123 NH-GQITAKDVQEALRDDTIMVSLMFVNNETGDPLPIQEIGQLLRNHQAVFHVDAVQVFS 181
           +  G+++ +DV++ALRDDTI+V+LM+ NNE G   PI EIG LL++HQAV H DAVQ F
Sbjct: 123 DQTGRVSVEDVRQALRDDTIIATTLMYGNNEVGTIQPIAEIGALLQDHQAVLHTDAVQARG 182

Query: 182 KMELDPHSLGIDFLAASAHKFHGPKGVGILYCAPH-HFDSLLHGGDQEKKRRASTENIIG 240
            + ++   L +D L+ SAHK +GPKGVG+LY          L+GG+QE K+RA TEN+
Sbjct: 183 AISIELDHLPVDMLSVSAHKINGPKGVGLLYVRDGIVLKPALYGGEQRKKRAGTENVAA 242

Query: 241 IAGMSQALTDATTNTLKNWTHISQLRTTFLDAISD--LDFYLNNGQDC-LPHVLNIGPFG 297
           I G ++A+  A  N  +          TF D      + F +N  Q    LPH+ N+ PFG
Sbjct: 243 IIGFAKAVEIAIANREERQKAYPDYQQTFFDQFQQEGVQFVMNGHQTNRLPHIFNVSPFG 302

Query: 298 QNNGLLLTQLDLAGFAVSTGSACTAGTVEPSHVLTSLYGANSPRLNESIRISPSELNTQE 357
           +   LL  LDLAG A S+GSACTAG++EPSHVL +++G++S  +  +R SF  NT+E
Sbjct: 303 VHVEALLVNLDLAGIAASSGSACTAGSIEPSHVLVAMHGSDSELVTSGVRPSPGLGNTKS 362

Query: 358 EILELAKTLRKII 370
           +  AK   KI+
Sbjct: 363 HVQWAAKETAKIV 375
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5271> which encodes the amino acid sequence <SEQ ID 5272>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1477(Affirmative) < succ>
```

```
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 235/370 (63%), Positives = 285/370 (76%)

Query: 2   MIYLDNAATTALTPSVIEKMTNVMTSNYGNPSSIHTFGRQANQLLRECRQIIAEYLNVNS 61
            M Y DNAATT L+P+VI  MT  M  N+GNPSSIH +GR+AN++LRECRQ IA  L  +
Sbjct: 1   MTYPDNAATTFLSPNVIRAMTAAMQDNFGNPSSIHPYGRRANKILRECRQAIARNLGASE 60

Query: 62  RRIIFTSGGTESNNTAIKGYALANQLKGKHIITSHIEHHSVLHTMTYLSERHGFDITYLK 121
            ++II TSGGTESNN AIKGYALA+Q KGKH+IT+ IEHHSVLHTM YL ERPGF++TYL
Sbjct: 61  QQIIVTSGGTESNNMAIKGYALAHQAKGKHLITTTIEHHSVLHTMAYLEERPGFEVTYLP 120

Query: 122 PVNHGQITAKDVQEALRDDTIMVSLMFVNNETGDFLPICEIGQLLRNHQAVFHVDAVQVFS 181
              +GQI    D+++ALRDDTI+VS+MYANNETGDLLPIKDIGNLLKDHQAAFHVDAVQAVG 180

Query: 182 KMKLDPHSLGIDFLAASAHKFHGPKGVGILYCAPHHFDSLLHGGDQEEKRRASTENIIGI 241
            K+++ P  LGIDFL+ASAHKFHGPKG G LY      D LLHGGDQE KRRASTEN++GI
Sbjct: 181 KLKIIPSELGIDFLSASAHKFHGPKGCGFLYSNGQPIDPLLHGGDQEGKRRASTENMLGI 240

Query: 242 AGMSQALTDATTNTLKNWTHISQLRTTFLDAISDLDFYLANGQDCLPNVLNIGFPGQNNG 301
             GM+QALTDA T   +-  HI  LR    +  L +Y+N G   LPHVLNIGF G  N
Sbjct: 241 IGMAQALTDAMTCLDQSTDHIISLRHHLISLLESDFYYINQGTHYLPHVLNIGFLGYQNT 300

Query: 302 LLLTQLDLAGFAVSTGSACTAGTVEPSHVLTSLYGANSPRLNESIRISFSELNTQEEILE 361
            +LLTQLDLAG AVSTGSACTAG V PSHVL + YG +S RL ESIRISFS+ N+ E++ +
Sbjct: 301 ILLTQLDLAGIAVSTGSACTAGAVNPSHVLAAYYGDDSSRLKESIRISFSDQNSIEDVNQ 360

Query: 362 LAKTLRKIIG 371
            LA+TL+ I+G
Sbjct: 361 LAQTLKNILG 370
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1698

A DNA sequence (GBSx1802) was identified in *S.agalactiae* <SEQ ID 5273> which encodes the amino acid sequence <SEQ ID 5274>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2753(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12416 GB:Z99107 ydiH [Bacillus subtilis]
 Identities = 96/202 (47%), Positives = 140/202 (68%), Gaps = 4/202 (1%)

Query: 7   IPKATAKRLSLYYRIFKRFNTDGIKKASSKQTADAIGIDSATVRRDFSYPGELGRKGFGY 66
            IP+ATAKRL LYYR  K  +  G ++ SS +++DA+ +DSAT+RRDFSYFG LG++G+GY
Sbjct: 8   IPQATAKRLPLYYRFLVNLHASGKQRVSSARLSDAVKVDSATRRDFSYFGALGKKGYGY 67

Query: 67  DVKKLMNFFAEILNDHSTTNVMLVGCGNIGRALLHYRPHDRNQMQISMAFDLDSNDLVGK 126
            +V  L++FF + L+    T+V+L+G GN+G A LHY F   N  +ISMAFD++ + + +
Sbjct: 68  NVDYLLSFFRKTLDQDEMTDVILIGVGNLGTAFLHYNFTKNNNTKISMAFDINESKI--G 125

Query: 127 TTRDGIFVYGISTINDHLIDSDIETAILTVPSTEAQEVADILVKAGIKGILSFSPVHLTL 186
            T  G+PVY +  + H+ D  + AILTVP+  AQ + D LV  GIKGIL+F+P  L +
```

```
Sbjct: 126 TEVGGVPVYNLDDLEQHVKDESV--AILTVPAVAAQSITDRLVALGIKGILNPTPARLNV 183

Query: 187 PKDIIVQYVDLTSELQTLLYFM 208
           P+ I + ++DL  ELQ+L+YF+
Sbjct: 184 PEHIRIHHIDLAVELQSLVYFL 205
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5275> which encodes the amino acid sequence <SEQ ID 5276>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2313(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 167/210 (79%), Positives = 189/210 (89%)

Query: 1   MIMDKSIPKATAKRLSLYYRIPKRFNTDGIEKASSKQIADALGIDSATVRRDFSYFGELG 60
           +++DKSIPKATAKRLSLYYRIPKRF+ D +EKASSKQIADA+GIDSATVRRDFSYFGELG
Sbjct: 1   VVIDKSIPKATAKRLSLYYRIPKRFHADQVEKASSKQIADAMGIDSATVRRDFSYFGELG 60

Query: 61  RRGPGYDVKKLMNFFAELLNDHSTTNVMLVGCGNIGRALLHYRFHDRNKMQISMAFDLDS 120
           RRGPGYDV KLMNFFA++LNDHSTTNV+LVGCGNIGRALLHYRFHDRNKMQI+M FD D
Sbjct: 61  RRGPGYDVTKLMNFFADLLNDHSTTNVILVGCGNIGRALLHYRFHDRNKMQIAMGFDTDD 120

Query: 121 NDLVGKTTEDGIPVYGISTINDHLIDSDIETAILTVPSTEAQEVADILVKAGIKGILSFS 180
           N LVG  T D IPV+GIS++ + + ++DIETAILTVPS  AQEV D L++AGIKGILSF+
Sbjct: 121 NALVGTKTADNIPVHGISSVKERIANTDIETAILTVPSIHAQEVTDQLIEAGIKGILSFA 180

Query: 181 PVHLTLPKDIIVQYVDLTSELQTLLYFMNQ 210
           PVHL +PK +IVQ VDLTSELQTLLYFMNQ
Sbjct: 181 PVHLQVPKGVIVQSVDLTSELQTLLYFMNQ 210
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1699

A DNA sequence (GBSx1803) was identified in *S.agalactiae* <SEQ ID 5277> which encodes the amino acid sequence <SEQ ID 5278>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2966(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9847> which encodes amino acid sequence <SEQ ID 9848> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14764 GB:Z99118 similar to DNA repair protein [Bacillus subtilis]
 Identities = 90/210 (42%), Positives = 136/210 (63%)

Query: 24  PRERLVDLGADRLSNQELLAILLRTGIKEKPVLEISTQILENISSLADFGQLSLQELQSI 83
           PRERL+ +GA+ L+N ELLAILLRTG K + VL++S ++L +    L    + S++EL SI
Sbjct: 19  PRERLLKVGAENLANEHLLAILLRTGTKHESVLDLSNRLLRSFDGLRLLKEASVEELSSI 78
```

```
Query: 84  KGIGQVKSVEIKAMLELAKRIHKAEYDRKEQILSSEQLARKMMLELGDKKQEHLVAIYMD 143
           GIG VK+++I A +EL  RIHK   +    I S E  A  +M ++     QEH V +Y++
Sbjct: 79  PGIGMVKAIQILAAVELGSRIHKLANEEHFVIRSPEDGANLVMEDMRFLTQEHFVCLYLN 138

Query: 144 TQNRIIEQRTIFIGTVRRSVAEPREILHYACKNMATSLIIIHNHPSGSPKPSESDLSFTK 203
           T+N++I +RT+FIG++  S+  PRE+   A K  A S I +HNHPSG P PS  D+  T+
Sbjct: 139 TKNQVIHKRTVFIGSLNNSSIVHPREVFKEAFKRSAASFICVHNHPSGDPTPSREDIEVTR 198

Query: 204 KIKRSCDHLGIVCLDHIIVGKNKYYSFREE 233
           ++    + +GI  LDH+++G  K+ S +E+
Sbjct: 199 RLFESCGNLIGIELLDHLVIGDKKFVSLKEK 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5279> which encodes the amino acid sequence <SEQ ID 5280>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3307(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 145/225 (64%), Positives = 182/225 (80%)

Query: 12  MYHIELKKEALLPRERLVDLGADRLSNQELLAILLRTGIKEKPVLEISTQILENISSLAD 71
           MY I+      +PRERL+ LGA+ LSNQELLAILLRTG KEK VLE+S+ +L ++ SLAD
Sbjct: 1   MYSIKCDDNKAMPRERLMRLGARSLSNQELLAILLRTGNKEKHVLELSSYLLSHLDSLAD 60

Query: 72  PGQLSLQELQGIKGIGQVKSVEIKAMLELAKRIHKAEYDRKEQILSSEQLARKMMLELGD 131
           P ++SLQELQ + GIG+VK++EIKAM+EL  RI   +    +L+S Q+A KMM  LGD
Sbjct: 61  PKMNSLQELQHLAGIGKVKAIEIKAMIELVSRILATDKTLVTDSVLTSVQVAEKMMAALGD 120

Query: 132 KKQEHLVAIYMDTQNRIIEQRTIFIGTVRRSVAEPREILHYACKNMATSLIIIHNHPSGS 191
           KKQEHLV +Y+D QNRIIE++TIFIGTVRRS+AEPREIL+YACKNMATSLI+HNHPSG+
Sbjct: 121 KKQEHLVVLYLDNQNRIIEKTIFIGTVRRSLAEPREILYYACKNMATSLIVIHNHPSGN 180

Query: 192 PKPSESDLSFTKKIKRSCDHLGIVCLDHIIVGKNKYYSFREEADI 236
            +PS +D  FT+KIKRSC+ LGI+CLDHIIV    YYSFRE++ +
Sbjct: 181 IEPSSNDYCFTEKIKRSCEDLGIICLDHIIVSYKDYYSFREKSTL 225
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1700

A DNA sequence (GBSx1804) was identified in *S.agalactiae* <SEQ ID 5281> which encodes the amino acid sequence <SEQ ID 5282>. This protein is predicted to be a permease. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.86   Transmembrane  258 - 274 ( 255 - 290)
   INTEGRAL    Likelihood = -7.32   Transmembrane   89 - 105 (  79 - 109)
   INTEGRAL    Likelihood = -4.88   Transmembrane  176 - 192 ( 170 - 194)
   INTEGRAL    Likelihood = -4.78   Transmembrane  339 - 355 ( 326 - 359)
   INTEGRAL    Likelihood = -4.57   Transmembrane  237 - 253 ( 236 - 257)
   INTEGRAL    Likelihood = -3.98   Transmembrane   39 -  55 (  38 -  59)
   INTEGRAL    Likelihood = -3.40   Transmembrane  292 - 308 ( 282 - 308)
   INTEGRAL    Likelihood = -1.38   Transmembrane  317 - 333 ( 317 - 333)
   INTEGRAL    Likelihood = -0.27   Transmembrane    8 -  24 (   8 -  24)
```

```
----- Final Results -----
                bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC05771 GB:AF051356 putative permease [Streptococcus mutans]
 Identities = 88/366 (24%), Positives = 175/366 (47%), Gaps = 27/366 (7%)

Query: 3   FEKRQVYYVVITFAICYAIQAYW---GAVSNILTTLHIKAIF-PFLMGAGIAYIINIVMSV 58
            F+  ++++  +   +   I   W   G++ N   ++ K  F PFL+G  + YI N +++
Sbjct: 2   FKSSKLFFWTVEILLVTLILFIWRQMGSIPNPFFSVAKTFFLPFLLGGPLYYITNPIVTF 61

Query: 59  YERLYIKLFKGSRLLMAIKKSVSMILSYATFIGLIVWLFSIVIPDLISSLSSLLVIDTGA 118
             E  +             IKR   + L +A  + L+V+  + +IP+LI+ L+ L+
Sbjct: 62  LENRF-----------KIKRIWGITLIFAVLLSLLVFSITSLIPNLINQLTDLISASQNI 110

Query: 119 LAKLVNNLNFNKQISEVLNYMGTDKDLVSTLSGYSQQILKQVLSVLTNLLTSVSSIAATL 178
              L +  NE K      N       D+   L  ++   + +VL ++  SVSSI    +
Sbjct: 111 YVGLQDLFNEWKSNPAFKNI-----DIPVLLKQFNLSYVDILTNVLDSVTVSVSSIVYMI 165

Query: 179 LNVFVSPIFS----ITVLANKEQLGRQFNLLIDTYLGSTGKTFHYVRHILHGRPHGFFVS 234
             N  +  + +     Y+L +K+ L      +L  T L +       + + +++     +
Sbjct: 166 TNTVMILVLTPVILFYLLKDKDGL---MPMLDRTILKNDRENISQLLNQMNKTISRYISG 222

Query: 235 QTLEAMILGSLTVIGMLIFQFPYALTVGVLVAFTALIPVVGAYIGVTIGFILIATESLTE 294
              ++A  +    +IG  I    YA    ++  T +IP VG Y+G+T   +          +
Sbjct: 223 VAIDAAFIFVFALIGYQIMSVQYAFLFALVAGITNVIPYVGPYLGLTPVVLAYVVSDPKK 282

Query: 295 AFLFVLFLILLQQFEGNVIYPKVVGGSIGLPSMMVLMAITIGGALWGILGMLLAVPVAAT 354
             + +++++ LQQ +GN++YP+VVG +  +  + +++ + +GG + G++GMLLAVP+A   A
Sbjct: 283 MIIAIIYIMTLQQIDGNIVYPRVVGSTMKIHPLTIMVLLVLGGNIAGLVGMLVAVPAYAI 342

Query: 355 IYQIVK 360
            I +IVK
Sbjct: 343 IKEIVK 348
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5283> which encodes the amino acid sequence <SEQ ID 5284>. Analysis of this protein sequence reveals the following:

```
      Possible site: 55
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.70   Transmembrane   87 - 103 (  83 - 116)
   INTEGRAL    Likelihood = -7.27   Transmembrane  178 - 194 ( 166 - 202)
   INTEGRAL    Likelihood = -6.74   Transmembrane  278 - 294 ( 256 - 297)
   INTEGRAL    Likelihood = -5.41   Transmembrane  299 - 315 ( 295 - 321)
   INTEGRAL    Likelihood = -4.46   Transmembrane   14 -  30 (  13 -  32)
   INTEGRAL    Likelihood = -3.56   Transmembrane  340 - 356 ( 333 - 366)
   INTEGRAL    Likelihood = -3.35   Transmembrane  258 - 274 ( 256 - 277)

----- Final Results -----
                bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC05771 GB:AF051356 putative permease [Streptococcus mutans]
 Identities = 87/373 (23%), Positives = 168/373 (44%), Gaps = 41/373 (10%)

Query: 10  FEKKQVFYLVLTFILCYGILANWRNGTAIVITTYKTS----LPFFYGAAGAYIVNIVMSA 65
            F+  ++F+  +  +L   IL  WR   +I    +  +    LPF  G    YI N +++
Sbjct: 2   FKSSKLFFWTVEILLVTLILFIWRQMGSIPNPFFSVAKTFFLPFLLGGPLYYITNPIVTF 61

Query: 66  YEKVYVYIFKDWSHVLKVKRGICLLLAYLTFFILYTWIISIVIPDLITSISTLTKFDT-- 123
```

```
            E  +           K+KR    + L +      L+ + I+ +IP+LI  ++ L
Sbjct: 62  LENRF-----------KIKRIWGITLIFAVLLSLLVFSITSLIPNLINQLTDLISASQNI 110

Query: 124 -ITIQEVVNNLEHNKLLARTIQYIGGDGKLTETIANYSQQLLKQFLTVLTNILTSVTVIA 182
            + +Q++ N  + N                    I     +Q    ++ +LTN+L SVTV
Sbjct: 111 YVGLQDLPNEWKSNPAFKNI------------DIPVLLKQFNLSYVDILTNVLDSVTVSV 158

Query: 183 SAIINLFISFVFSL--------YVLASKEDLCRQGNTLVDTYTGKYAKRIHYLLELLHQR 234
            S+I+ +  + V  L        Y+L  K+ L       L  T       I  LL  +++
Sbjct: 159 SSIVYMITNTVMILVLTPVILFYLLKDKDGLMPM---LDRTILKNDRHNISQLLNQMNKT 215

Query: 235 FHGPFVSQTLEAMILGSLTASGMFILRLPFAGTIGVLVAFTALIPVIGASIGAAIGFLI 294
               +      ++A  +      G  I+ + +A     ++    T +IP +G  +G      +
Sbjct: 216 ISRYISGVAIDAAFIFVFALIGVQIMGVQVAPLFALVAGITNVIPVVGFYLGLTPVVLAY 275

Query: 295 MTQSMSQAIIFIIFLIILQQIEGNFIYPKVVGGSIGLPAMWVLMAITIGASLKGIVGMIY 354
            +      + II II+++ LQQI+GN +YP+VVG ++ +  + +++ + +G ++ G+VGM++
Sbjct: 276 VVSDPKKMIIAIIYIMTLQQIDGNIVYPRVVGSTMKIHPLTIMVLLVLGGNIAGLVGMLV 335

Query: 355 AVPLAATLYQVIK 367
            AVP  A + +++K
Sbjct: 336 AVPAYAIIKEIVK 348
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 218/370 (58%), Positives = 291/370 (77%)

Query: 1   MKFSKRQVYYVVTFAICYAIQAYWGAVSNILITTLHKAIFPFLMGAGIAYIINIVMSVYE 60
            MKFEK+QV+Y+V+TF +CY I A W    + I+TT++K   PF  GA  AYI+NIVMS YE
Sbjct: 8   MKFEKKQVFYLVLTFILCYGILANWRNGTAIVTTIYKTSLPFFYGAAGAYIVNIVMSAYE 67

Query: 61  RLYIKLFKGSRLLMAIKRSVSMILSYATFIGLIVWLPSIVIPDLISSLSSLLVIDTGALA 120
            ++Y+ +FK    ++ +KR + ++L+Y TF  LI W+ SIVIPDLI+S+S+L   DT  +
Sbjct: 68  KVYVYIFKDWSHVLKVXRGICLLLAYLTFFILTTWIISIVIPDLITSISTLTKFDTITIQ 127

Query: 121 KLVNNLNENKQISEVLNYMGTDKDLVSTLGGYSQQILKQVLSVLTNLLTSVSSIAATLLN 180
            ++VNNL  NK ++  + Y+G D  L  T++ YSQQ+LKQ L+VLTN+LTSV+ IA+ ++N
Sbjct: 128 EVVNNLEHNKLLARTIQYIGGDGKLTETIANYSQQLLKQFLTVLTNILTSVTVIASAIIN 187

Query: 181 VFVSFIPSIYVLANKEQLGRQFNLLIDTYLGSTGKTFHYVRHILHQRFHGFFVSQTLEAM 240
            +F+SF+PS+YVLA+KE L RQ N L+DTY G   K  HY+  +LHQRFHGFFVSQTLEAM
Sbjct: 188 LFISFVPSLYVLASKEDLCRQGNTLVDTYTGKYAKRIHYLLELHQRFHGFFVSQTLEAM 247

Query: 241 ILGSLTVIGMLIPQFPYALTVGVLVAFTALIPVVGAYIGVTIGFILIATESLTEAFLFVL 300
            ILGSLT  GM I + P+A T+GVLVAFTALIPV+GA IG  IGFILI T+S+++A +F++
Sbjct: 248 ILGSLTASGMFILRLPFAGTIGVLVAFTALIPVIGASIGAAIGFILIMTQSMSQAIIFII 307

Query: 301 FLILLQQFEGNVIYPKVVGGSIGLPSMWVLMAITIGGALWGILGMLLAVPVAATIVQIVK 360
            FLI+LQQ EGN IYPKVVGGSIGLP+MWVLMAITIG +L GI+GM++AVP+AAT+VQ++K
Sbjct: 308 FLIILQQIEGNFIYPKVVGGSIGLPAMWVLMAITIGASLKGIVGMIIAVPLAATLYQVIK 367

Query: 361 DHIIKRQTLR 370
            D+I KRQ ++
Sbjct: 368 DNIQKRQAIQ 377
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1701

A DNA sequence (GBSx1805) was identified in *S.agalactiae* <SEQ ID 5285> which eneodes the amino acid sequence <SEQ ID 5286>. Analysis of this protein sequence reveals the following:

```
  Possible site: 18
  >>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
                bacterial cytoplasm --- Certainty=0.1081(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9849> which encodes amino acid sequence <SEQ ID 9850> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA69226 GB:U29579 6-phospho-beta-glucosidase [Escherichia coli]
 Identities = 290/478 (60%), Positives = 369/478 (76%), Gaps = 2/478 (0%)

Query: 2   MVKQVFPKGFLWGGATAANQCEGAYNVDGRGLANVDVVPTGEDRFAIISGQKDMPDFEEG 61
            M   VFP+ FLWGGA AANQ EGA+     +GL  VD++P GE R A+  G +K F    +
Sbjct: 1   MKMSVFPESFLWGGALAANQSEGAFREGDKGLTTVDMIPHGEHRMAVKLGLERKFQLRDD 60

Query: 62  YFYPAKESIDFYHHYKEDLALLAEMGFKTYRMSIAWTRIFPKGDELYPNEAGLQFYENIF 121
             FYP+ E+ DFYH YKED+AL+AEMGFK +R SIAW+R+FP+GDE+ PN+ G+ FY ++F
Sbjct: 61  EFYPSHEATDFYHRYKEDIALMAEMGFKVFRTSIAWSRLFPQGDEITPNQQGIAFYRSVF 120

Query: 122 KECRKYGIEPLVTITHFDCPIYLIKHYGGWRSRKMIGFVERLVRALFTRFKGLAVCYWLTF 181
            +EC+KYGIEPLVT+ HFD F++L+  YG WR+RK++ F+ R  R  F  F GLAVCYWLTF
Sbjct: 121 EECKKYGIEPLVTLCHFDVPMHLVTEYGSWRNRKLVEFFSRYARTCFEAFDGLAVCYWLTF 180

Query: 182 NEINMILHAPFWGAGLYFEDGENQEQIKYQAAHHELVASAIAVKIAHEVDPNNQIGCMLA 241
            NEIN++LH+PF GAGL FE+GENQ+Q+KYQAAHH+LVASA+A KIAHEV+P NQ+GCMLA
Sbjct: 181 NEINIMLHSPFSGAGLVFEEGENQDQVKYQAAHHQLVASALATKIAHEVNPQNQVGCMLA 240

Query: 242 AGQYYPNTCHPQDYWASMQKNRENYFFIDVQARGKYPNYAKKHFEHLGISIQMTAEDLAL 301
             G +YP +C P+D WA+++K+REN FFIDVQARG YP Y+ + F   G++I     D  +
Sbjct: 241 GGNFYPYSCKPEDVWAALEKDRENLFFIDVQARGTYPAYSARVFREKGVTINKAPGDDEI 300

Query: 302 LRDYTVDFISFSYYSSRVASGNPTVSEQVQENIFASLVNPYVLKSSEWGWQIDPLGLRITL 361
            L++ TVDF+SFSYY+SR AS     +   N+  SL+NPYL+ S+WGW IDPLGLRIT+
Sbjct: 301 LXN-TVDFVSFSYYASRCASAENNANNSSAANVVKSLRNPYLQVSDWGWGIDPLGLRITM 359

Query: 362 NAIWDRYQKPMFIVENGLGAVDIPDENGVVEDDYRIDYLRQHIAAMRDAITVDGVNLIGY 421
            N ++DRYQKP+F+VENGLGA D    NG + DDYRI YLR+HI AM +AI  DG+ L+GY
Sbjct: 360 NMMYDRYQKPLFLVENGLGAKDEFAANGEINDDYRISYLREHIRAMGEAI-ADGIPLMGY 418

Query: 422 TTWGCIDLVSAGTGEMEKRYGFIYVDRNNKGEGTLKRYKKKSFYWYKKVIASNGSQIE 479
            TTWGCIDLVSA TGEM KRYGF++VDR++ G GTL R +KKSF+WYKKVIASNG  +E
Sbjct: 419 TTWGCIDLVSASTGEMSKRYGFVFVDRDDAGNGTLTRTKKKSFWWYKKVIASNGEDLE 476
```

There is also homology to SEQ ID 5288.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1702

A DNA sequence (GBSx1806) was identified in *S.agalactiae* <SEQ ID 5289> which encodes the amino acid sequence <SEQ ID 5290>. This protein is predicted to be platelet-activating factor acetylhydrolase isoform Ib beta subunit, pu. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.5323(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC27974 GB:AF016048 platelet-activating factor acetylhydrolase
          alpha 2 subunit [Rattus norvegicus]
 Identities = 43/177 (24%), Positives = 84/177 (47%), Gaps = 9/177 (5%)

Query: 28  QEGAIVFTGDSIVEF---FPLKKHLGRDYPLVNRGVAGSDTYWLLENLRTQVWELLPSKV 84
           +E  ++F GDS+V+    + + + L       +N G+ G  T  +L  L+    E +  KV
Sbjct: 38  KEPDVLFVGDSMVQLMQQYETWRELFSPLHALNFGIGGDTTRHVLWRLKNGKLENIKPKV 97

Query: 85  FIL-IGTNDIGLGHSQSEIIANITDIIAEIRAESYMTEINILLSVLPVSEEDDYIERVKVR 143
           ++ +GTN+     ++  E+   I  I+  I          +I +L +LP   E+ + + +
Sbjct: 98  IVWNVGTNNHE---NTAEEVAGGIEAIVQLINTRQPQAKIIVLGLLPRGEKPNPLRQKNAK 155

Query: 144 NNQTIKALNKTLSVISGINYIELYDLLVDEKGQLASSFTKDGLHLTDQAYAKISETI 200
             NQ +K    +L  ++ +  +++     V   G ++       D LHLT   YAKI + +
Sbjct: 156 VNQLLKV---SLPKLANVQLLDIDGGFVHSDGAISCHDMPDFLHLTGGGYAKICKPL 209
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5291> which encodes the amino acid sequence <SEQ ID 5292>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5979(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 92/204 (45%), Positives = 133/204 (65%)

Query: 1   MLEVIDKALADYQMKREQPFEINMQTVQEGAIVFTGDSIVEFFPLKKHLGRDYPLVNRGV 60
           MLE++ + LR YQ ++    +   NQ   +G IVF GDS++EFFPLKK  G   P++NRG+
Sbjct: 1   MLEIVSEELRHYQEQKLIEYRNKNQLAPKGGIVFAGDSLIEFFPLKKAFGSCLPIINRGI 60

Query: 61  AGSDTYWLLENLRTQVWELLPSKVPILIGTNDIGLGHSQSEIIANITDIIAEIRAESYMT 120
           AG D+ WLL +   Q+ +L P  +P+LIG NDIGLG+ +   I+  I ++I++IR+     +
Sbjct: 61  AGIDSQWLLRHFSVQITDLRPKHIPLLIGCNDIGLGYDKCHIVKTIVELISQIRSHCVYS 120

Query: 121 EINILSVLPVSEEDDYIERVKVRNNQTIKALNKTLSVISGINYIELYDLLVDEKGQLASS 180
           +I +LS+LPVS    Y + VK+R N  I A+NK L+ I  + +I L   L DEKG L+S
Sbjct: 121 QIYLLSLLPVSNNPRYQKTVKIRTNAMIDAINKDLAMIPTVEPINLNTCLKDEKGGLSDE 180

Query: 181 FTKDGLHLTDQAYAKISETIKLYL 204
            T DGLHL   AYAK+++E IK Y+
Sbjct: 181 NTLDGLHLNFPAYAKLIKSYI 204
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1703

A DNA sequence (GBSx1807) was identified in *S.agalactiae* <SEQ ID 5293> which encodes the amino acid sequence <SEQ ID 5294>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5226(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9851> which encodes amino acid sequence <SEQ ID 9852> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA35556 GB:D90723 Hypothetical 30.2 kd protein in idh-deoR
          intergenic region. [Escherichia coli]
 Identities = 104/265 (39%), Positives = 154/265 (57%), Gaps = 4/265 (1%)

Query: 2   IKLIATDMDGTFLRSDKTYDKARFSSLIQTLMEKYDIKFVAASGNLYDQLLLNFLEYPNRI 61
           IKLIA DMDGTFL   KTY++ RF +     M+   I+FV ASGN Y QL+  F E  N I
Sbjct: 4   IKLIAVDMDGTFLSDQKTYNRERFMAQYQQMKAQGIRFVVASGNQYYQLISFFPEIANEI 63

Query: 62  AYVAENGGRVIDQDGTLLKETYLSNDTVAAVLSYLYQNYPETLISLSGEKRSYLERRTPI 121
           A+VAENGG V+ + G +    LS D  A V+ +L    PE  I   G+  +Y  ++
Sbjct: 64  AFVAENGGWVVSE-GKDVFNGELSKDAFATVVSHLLTR-PEVEIIACGKNSAYTLKKYDD 121

Query: 122 NRRTTELEYYMPNFIYKDHLLPLDDDRYPQMTLWVNENLVSEMLLDISEHFKNHHIRLTSS 181
            +T  E Y      Y D+   L+D  +F+  L +++ L+  ++    + E   +  + +
Sbjct: 122 AMKTVAEMYYHRLEYVDNFDNLEDI-FFKFGLNLSDELIPQVQKALHEAIGDIMVSV-HT 179

Query: 182 GFGCIDVLPADVNKADGIAILLEKWGLKQDQVMVFGDGGNDVEMLRAANISYAMSNAPEE 241
           G G ID++    V+KA+G+  L + WG+    +V+VFGDGGND+EMLR A  S+AM NA
Sbjct: 180 GNGSIDLIIPGVHKANGLRQLQKLWGIDDSEVVVFGDGGNDIEMLRQAGFSFAMENAGSA 239

Query: 242 IKAIAKVQTVSNDQDGVLETIENFL 266
           + A AKV+  SN+++GVL+ I+  L
Sbjct: 240 VVAAAKYRAGSNNREGVLDVIDKVL 264
```

There is also homology to SEQ ID 1158.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1704

A DNA sequence (GBSx1808) was identified in *S.agalactiae* <SEQ ID 5295> which encodes the amino acid sequence <SEQ ID 5296>. This protein is predicted to be transcriptional regulator (AraC/XylSfamily). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4984(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF89977 GB:AF206272 transcriptional regulator [Streptococcus mutans]
 Identities = 195/287 (67%), Positives = 237/287 (81%)

Query: 5   DNLLSHNLEDNNRHLLPYEHMHTEVRNGYPDILFHWHPELEISYVHEGTARYHIDYDFPNS 64
            D    H +  +  LLPY+    T + NGYPD LFHWHPELEISY++EGTA+YHIDYD+FNS
Sbjct: 10  DENFKHEINFDNDLLPYKIYQTTIANGYPDTLFHWHPELEISYIYEGTAQYHIDYDYPNS 69

Query: 65  QSGDIILIRPNGMHSIHPIENKEHITDSIKWHLDLIGYSIVDQVSLRYLQPLQTSSFKFI 124
           Q+ DIIL+RPNGMHSIHPI+NK      ++ FHLDL+GYS++DQ+SLRYLQPLQ S+FK +
Sbjct: 70  QTDDIILVRPNGMHSIHPIKNKMQKAQTLLFHLDLVGYSLLDQISLRYLQPLQNSTFKLV 129

Query: 125 QCIKPSMTGYNDIKNCLEDIFNISKENRHFELLLKAKLNELLYLYHQYVIKGHTDDT 184
            CIKP N GY DIKNCLF IF+I + + RHFELLLKAKL EL+YLLY+HQYV++KH+DD
Sbjct: 130 PCIKPDNLGYQDIKNCLFAIFDIYQRQGRHFELLLKAKLQELIYLLYPHQYVLRKHSDDN 189
```

```
Query: 185 YRKNERIRDLIDYINNNYQQNLTIEFLADYMGYSKTHFMTVFKQHTGTSCTEFIIQVRLN 244
           YRKNE+IR+LIDYI+ +YQ+ L+I  LAD +GYSKTHFMTVFKQHTGTSCT+FIIQ RL+
Sbjct: 190 YRKNEKIRELIDYIHQHYQEKLSIISLADIIGYSKTHFMTVFKQHTGTSCTDFIIQFRLS 249

Query: 245 KASEHLINSTTAIIDIANSVGFNNLSNFNRQFKRYYHTTPRQYRKQF 291
           KA + L+NS    I+++A+ VGF NLSNFNRQFKRYY  TP QYRKQF
Sbjct: 250 KACDLLVNSIKPILEVASEVGFTNLSNFNRQFKRYYQITPSQYRKQF 296
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5297> which encodes the amino acid sequence <SEQ ID 5298>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 43/169 (25%), Positives = 83/169 (48%), Gaps = 16/169 (9%)

Query: 136 DIKNCLFDIFNISKENRHFELLLKAKLNELLYLLYYHQYV------IKKHTTDTYRKN- 188
           D+K+  F +F+   + R F +L K     ++  ++  Q +      +KK  D T + N
Sbjct: 319 DVKHVSFLLFS---DIYRQFPILDKWTYLSMVKTIHDSQSIDCILRELKKVLDVTNQNNS 375

Query: 189 ------ERIRDLIDYINNNYQQNLTIEFLADYMGYSKTHFMTVFKQHTGTSCTEFIIQVR 242
                 + + + ID I   Y Q LT++ +AD +  +  +    FK  T  S T+++  VR
Sbjct: 376 PEKRYSDLVSETIDCIRKEYHQELTLKAIADRLHVNGVYLGQCFKNETERSFTQYLNHVR 435

Query: 243 LNKASEHLINSTTAIIDIANSVGFNNLSNFNRQFKRYYHTTPRQYRKQF 291
           + KA + L+ +  +I +IA   G+N    F + FK+    +P+++R ++
Sbjct: 436 IQKAQQLLLYTNQSINEIAYETGYNTNHYFIKMFKKLNGLSPKEFRDRY 484
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1705

A DNA sequence (GBSx1809) was identified in *S.agalactiae* <SEQ ID 5299> which encodes the amino acid sequence <SEQ ID 5300>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3705(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1706

A DNA sequence (GBSx1810) was identified in *S.agalactiae* <SEQ ID 5301> which encodes the amino acid sequence <SEQ ID 5302>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.25   Transmembrane   59 -  75 (  56 -  82)
   INTEGRAL    Likelihood = -7.48   Transmembrane   23 -  39 (  12 -  41)
   INTEGRAL    Likelihood = -6.64   Transmembrane  231 - 247 ( 225 - 255)
   INTEGRAL    Likelihood = -5.15   Transmembrane  335 - 351 ( 333 - 355)
   INTEGRAL    Likelihood = -4.19   Transmembrane  309 - 325 ( 305 - 327)
   INTEGRAL    Likelihood = -4.14   Transmembrane  272 - 288 ( 268 - 292)
   INTEGRAL    Likelihood = -4.04   Transmembrane  402 - 418 ( 400 - 419)
   INTEGRAL    Likelihood = -3.88   Transmembrane  191 - 207 ( 190 - 208)
   INTEGRAL    Likelihood = -2.71   Transmembrane  365 - 381 ( 364 - 381)
   INTEGRAL    Likelihood = -1.86   Transmembrane  165 - 181 ( 164 - 182)

----- Final Results -----
            bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF96429 GB:AE004383 conserved hypothetical protein [Vibrio cholerae]
 Identities = 142/443 (32%), Positives = 241/443 (54%), Gaps = 20/443 (4%)

Query: 6   NEPQFSLESILGFVWRGIVVGLIAGFVVSIPRLAIEKIFLVVMELYKS--AHYQPIILLS 63
           N+F    ++    ++  ++VG++AG V + F  A+  +    +  KS    + P+ L +
Sbjct: 21  NQFLSKDKTPPSVLFLSLLVGILAGLVGTYFEQAVHLVSETRTDWLKSEIGSPLPLWLAA 80

Query: 64  ITVTSIIAAVIIGFFI--KSDPDIKGSGIPHVEGELKGMLSPDWFSIVWKKFTAGILAIS 121
             +++ +A   IG+F+  +  P+  GSGIP +EG + GM    W+ ++  KF  G+ A+
Sbjct: 81  FLISAFLA--FIGYFLVHRFAPEAAGSGIPEIEGAMDGMRPVRWMRVLPVKFPGGMGALG 138

Query: 122 SGLMLGREGPSIQLGAMIGKGIAQYLNASRMEKR-VLIASGAAAGLSAAFNAPIAGLLFV 180
            SG++LGREGP++Q+G   G+ I+       + R  L+A+GAA GL+AAFNAP+AG++FV
Sbjct: 139 SGMVLGREGPTVQMGSAVGRNISDIFRVKNEDTRHSLLAAGAAGCLAAAFNAPLAGIMFV 198

Query: 181 VEEIYHHFS-RLVWITALVASLV-ANFVSLNIFGLTPVLALPSELPSLNLNFVWIFLLWG 238
            +EE+  F    L+ + A++ S V AN V   I G   V+ +P +  +  L+   +FLL+G
Sbjct: 199 IEEMRPQFRYTLISVRAVIISAVAANIVPRVINGQDAVITMP-QYDAPELSTLGLFLLLG 257

Query: 239 LPLGILGFIYEWVIL----RFHVIYDYLGKLFHLPSHLVGILAVIFILPIGYYPPQLLGG 294
             G+ G ++ ++I      F   +  K + L  + G   + +L   Y P+L GG
Sbjct: 258 ALFGVFGVLFNYLITLAQDLFVKFHRNDRKRYLLTGSMIGGCFGLLLL----YVPELTGG 313

Query: 295 GNGLIVSLPRSNLSLMMLGLFFLIRFLWSMLSYSSGLPGGIFLPILALGSLAG-AFFAVG 353
           G  LI ++         +L L F+ R   ++L + SG PGGIF P+LALG+L G AF   +
Sbjct: 314 GISLIPTITNGGYGAGILLLLFVGRIFTTLLCFGSGAPGGIFAPMLALGTLFGYAFGLIA 373

Query: 354 MQYFGIISHQQISLFVVLGMAGYFGAISKADLTAMILVTEMVGDLKQLMAIGIVTMVSYI 413
            +F  ++  +P + GM   F A  +AP+T ++LV EM  +   ++ I ++ + I
Sbjct: 374 KWWFPELNIEP-GMFAIAGMGALFAATVRAPITGILLVIRMTNNYHLILPLITSLGAVI 432

Query: 414 VMDLLKGKPIYSLLAKMTFNPK 436
             LL G+PIY  +L +   N K
Sbjct: 433 FAQLLGGQPIYSQLLERTLKNQK 455
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5303> which encodes the amino acid sequence <SEQ ID 5304>. Analysis of this protein sequence reveals the following:

```
      Possible site: 31
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.68   Transmembrane   71 -  87 (  66 -  95)
   INTEGRAL    Likelihood = -9.45   Transmembrane   36 -  52 (  26 -  56)
```

```
INTEGRAL    Likelihood = -5.63   Transmembrane  346 - 362 ( 342 - 367)
INTEGRAL    Likelihood = -5.36   Transmembrane  376 - 392 ( 375 - 393)
INTEGRAL    Likelihood = -5.15   Transmembrane  413 - 429 ( 410 - 432)
INTEGRAL    Likelihood = -5.10   Transmembrane  321 - 337 ( 318 - 340)
INTEGRAL    Likelihood = -4.19   Transmembrane  203 - 219 ( 202 - 220)
INTEGRAL    Likelihood = -4.19   Transmembrane  244 - 260 ( 242 - 265)
INTEGRAL    Likelihood = -4.19   Transmembrane  284 - 300 ( 280 - 304)
INTEGRAL    Likelihood = -1.86   Transmembrane  177 - 193 ( 176 - 194)

----- Final Results -----
            bacterial membrane --- Certainty=0.5670(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF96429 GB:AE004383 conserved hypothetical protein [Vibrio cholerae]
 Identities = 144/442 (32%), Positives = 236/442 (52%), Gaps = 30/442 (6%)

Query: 18  NSFTFSNKSIIAYVWRGVVVGIIAGVIVSLFRLLIEVTADWVIEWYRYAHINSLLLPIL 77
           N+F   +K+  + ++  ++VGI+AG++ + F   + ++   +W + + I S L L +
Sbjct: 21  NQFLSKDKTPPSVLFLSLLVGILAGLVGTYFEQAVHLVSETRTDWLK-SEIGSFLPLWLA 79

Query: 78  SVSLLAVL-FVGFLV--KSDSDIKGSGIPHVEGELKGLMSPDWWSVLWKKFLGGIMAISM 134
           +  + A L F+G+ +  +   +  GSGIP +EG + G+    WW VL  KF GG+ A+
Sbjct: 80  AFLISAFLAFIGYFLVHRFAPEAAGSGIPEIEGAMDSMRPVRWWRVLPVKFFGGMGALGS 139

Query: 135 GPMLGREGPSIQLGAMSAKGLAKFLKSSRLEKR-VLIASGAAAGLSAAFNAPIAGLLFVV 193
           G +LGREGP++Q+G   + ++   +  + R  L+A+GAA GL+AAFNAP+AG++FV+
Sbjct: 140 GMVLGREGPTVQMGGAVGRMISDIFRVKNSDTRHSLLAAGAAGGLAAAFNAPLAGIMFVI 199

Query: 194 EEIYHHIPS-RILIWITALVASLV-ANFISLNIPGLKPVLAMSEAMPFLGLNQYWLLLLLGL 251
           EE+   F   LI + A++ S V AN +   I G   V+ M +       L+   L LLLG
Sbjct: 200 EEMRPQFRYTLISVRAVIISAVAANIVFKVINGQDAVITMPQ-YDAPELSTLGLFLLLGA 258

Query: 252 FLGCLGYLYEIVIL----------NFNKLYVILGSWLHLPDYFYGIIMVFLILPIGYYL 300
             G G L+  +I              N  K Y++ GS +        +G++++        Y+
Sbjct: 259 LFGVFGVLFNYLITLAQDLFVKFHRNDRKRYLLTGSMI---GGCFGLLLL--------YV 307

Query: 301 PQLLGGGHGLILLSLSNQQLFLMTIPFYFIIRFIVSMFSYSGGLPGGIFLPILTLGALAGL 360
           P+L GGG  LI ++N      +  F+ R  ++  +GSG PGGIF P+L LG L G
Sbjct: 308 PELTGGGISLIPTTTNGGYGAGILLLFVGRIFTTLLCFSSGAPGGIFAPMLALGTLFGY 367

Query: 361 LFGQIASQLGLLNQSFLSLFLILGMAGYFAAISKAPLTGMILVTEMVGDLKPLMAIAVVT 420
            FG IA            +F I GM   FAA  +AP+TG++LV EM  +   ++ + + +
Sbjct: 368 AFGLIAKMWFPELNIEPGMFAIAGMGALFAATVRAPITGILLVIEMTNNYHLILPLITS 427

Query: 421 FVSYLVMDLLNGQPIYRAMLDK 442
             + +   LL GQPIY  +L +
Sbjct: 428 LGAVIFAQLLGGQPIYSQLLHR 449
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 343/510 (67%), Positives = 410/510 (80%)

Query: 1   MENHKNEPQFSLKSILGFVWRGIVVGLIAGFVVSIFRLAIEKILFVVMRLYKSAHYQPII 60
           MENHKNEF FS +SI+ +VWRG+VVG+IAG +VS+FRL IE    V+E Y+ AH   ++
Sbjct: 13  MENHKNEFTFSNKSIIAYVWRGVVVGIIAGVIVSLFRLLIEVTADWVIEWYRYAHINSLL 72

Query: 61  LLSITVTSIIAAVIIGFFIKSDPDIKGSGIPHVEGELKGMLSPDWFSIVWKKFIAGILAI 120
           LL I   S++A + +GF +KSD DIKGSGIPHVEGELKG++SPDW+S++WKKF+ GI+AI
Sbjct: 73  LLPILSVSLLAVLFVGFLVKSDSDIKGSGIPHVEGELKGLMSPDWWSVLWKKFLGGIMAI 132

Query: 121 SSGLMLGREGPSIQLGAMTGKGIAQYLNASRMSKRVLIASGAAAGLSAAFNAPIAGLLFV 180
           S G MLGREGPSIQLGAM+ KG+A++L +SR+EKRVLIASGAAAGLSAAFNAPIAGLLFV
Sbjct: 133 SMGFMLGREGPSIQLGAMSAKGLAKFLKSSRLEKRVLIASGAAAGLSAAFNAPIAGLLFV 192

Query: 181 VEEIYHHFSRLVWITALVASLVANFVSLNIPGLTPVLALPSELPSLNLNFYWIFLLMGLF 240
```

```
              VEEIYHHFSRL+WITALVASLVANF+SLNIPGL PVLA+   +P L LN YW+ LL+GLF
Sbjct: 193 VEEIYHHFSRLIWITALVASLVANFISLNIPGLKPVLAWSEAMPFLGLNQYWLLLLLGLF 252

Query: 241 LGILGFIYEWVILRFHVIYDYLGKLPHLPSHLYGILAVIFILPIGYYFPQLLGGGNGLIV 300
           LG LG++YE VIL F+ +Y  LG   HLP + YGI+ V  ILPIGYY PQLLGGG+GLI+
Sbjct: 253 LGCLGYLYEIVILNFNKLYVILQSWLHLPDYFYGIIMVFLILPIGYYLPQLLGGGHGLIL 312

Query: 301 SLPRSNLSLMMLGLFFLIRFLWSMLSYSSGLPGGIFLPILALGSLAGAPFAVGMQYFGII 360
           SL   L LM +  +F+IRF+ SM SY SGLPGGIFLPILLGSALAGLFPGQIASQLGLL
Sbjct: 313 SLSNCQLPLMTIPFYFIRFIVSMPSYGSGLPGGIFLPILTLGALAGLLFPGQIASQLGLL 372

Query: 361 SHQQISLFVVLGMAGYFGAISKAPLTAMILVTRMVGDLKQLMAIGIVCMVSYIVMDLLKG 420
           +   +SLP++LGMAGYF AISKAPLT MILVTRMVGDLK LMAI +VT VSY+VMDLL G
Sbjct: 373 NQSPLSLFLILGMAGYFAAISKAPLTGMILVTRMVGDLKPLMAIAVVTPVSYLVMDLLNG 432

Query: 421 EPIYEAMLAKMTPNPKDKVMTPTLIELTVSDKISGKYVRDLELPENVLITTQIHHKTSAV 480
           +PIYEAML KM        ++ PTLIELTV DKI+GKYV++L+LPENVLITTQIHH+ S V
Sbjct: 433 QPIYEAMLDKMAMKHPTNLVEPTLIELTVGDKIAGKYVKELKLPENVLITTQIHHQKSQV 492

Query: 481 VSGNTILNAGDTIFLVVNESEIKEVREQLM 510
           VSGNT L +G TIFLVVNE++   VRE LM
Sbjct: 493 VSGNTRLLSGATIFLVVNEADTGFVREVLM 522
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1707

A DNA sequence (GBSx1811) was identified in *S.agalactiae* <SEQ ID 5305> which encodes the amino acid sequence <SEQ ID 5306>. This protein is predicted to be spermidine/putrescine-binding periplasmic protein precursor (potD-1). Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood = -9.02   Transmembrane   20 -  36 (  14 -  40)

----- Final Results -----
                bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8881> which encodes amino acid sequence <SEQ ID 8882> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 2
SRCFLG: 0
McG: Length of UR:   22
     Peak Value of UR:   4.16
     Net Charge of CR: 2
McG: Discrim Score:     18.94
GvH: Signal Score (-7.5): -3.29
     Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 1 value:  -9.02 threshold:  0.0
   INTEGRAL   Likelihood = -9.02   Transmembrane    7 -  23 (   1 -  27)
   PERIPHERAL  Likelihood =  6.05      170
 modified ALOM score:   2.30
icm1 HYPID: 7  CFP: 0.461

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF94581 GB:AE004221 spermidine/putrescine ABC transporter,
          periplasmic spermidine/putrescine-binding protein [Vibrio cholerae]
 Identities = 126/327 (38%), Positives = 196/327 (59%), Gaps = 2/327 (0%)

Query: 42  SSSTPNSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSNEAMHTKIKQGGTTYDIA 101
           +++     +L   YNW +YI   +L+ FTKETGI+V Y T++SNS+M+ K+K  G  YD+
Sbjct: 18  TNAMAKDQELYFYNWSEYIPSEVLEDFTKETGIKVIYSTYESNESMYAKLKTQGAGYDLV 77

Query: 102 VPSDYMIDKMIKENLLVKLDHSKIANWDAIGARFKNLSFDPGNKYSIPYFWGTVGIVYN- 160
           VPS Y + KM KE +L ++DHSK+++   +    + N  FDP NK+SIPY WG  GI  N
Sbjct: 78  VPSTYFVSKMRKEGMLQEIDHSKLSHFKDLDPNYLNKPFDPGNKFSIPYIWGATGIGINT 137

Query: 161 DQLVKTPPKNWDDLWRPEFRNKIMLVDSAREVIGVGLNSLGYGLNTKNISELKAASKKLD 220
           D L K   K+W DLW  ++  ++ML+D AREV  + L+ LGY  NT N  E+KAA ++L
Sbjct: 138 DMLDKKSLKNWGDLWDAKWAGQLMLMDDAREVFHIALSKLGYSPNTTNPKEIKAAYRELK 197

Query: 221 ALTPNVKAIVADEMKGYMIQGDAAIGVTFSGEAREMLDGNKHLHYVVPSEGSNLWFDNIV 280
            L PNV    +D     + G+ ++G+ ++G A     +  + P +G+  W D+I
Sbjct: 198 KLMPNVLVFNSDFPANFYLAGEVSLGMLWNGSAYMARQEGAPIQIIWPEKGTIPWMDSIS 257

Query: 281 IPKTVKHRKEAYAFINFWMEPKNNAQNAEYIGYATPNLKAKALLPADIKNDKAFYPDPDKT 340
           IP   K+ + A+  I+F++ P+NAA+ A  IGY TP   A  LLP +  ND + YPP
Sbjct: 258 IPAGAKNIEAAHKMIDFLRPENAAKIALEIGYPTPVKTAHDLLPKEPANDPSIYPPQSV 317

Query: 341 IDHLEVYNNLGQKWLGIYNDLYLQFKM 367
           ID+ E  + +G+  + +Y++ + + K+
Sbjct: 318 IDNGEWQDEVGEASV-LYDEYFQKLKV 343
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5307> which encodes the amino acid sequence <SEQ ID 5308>. Analysis of this protein sequence reveals the following:

```
      Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood = -8.44   Transmembrane    8 -  24 (   1 -  27)

----- Final Results -----
               bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC74207 GB:AE000212 spermidine/putrescine periplasmic transport
          protein [Escherichia coli]
 Identities = 134/342 (39%), Positives = 199/342 (58%), Gaps = 3/342 (0%)

Query: 17  ILTSLSFILQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDSNEAMYT 76
           +L + +  L   +     ++ L  YNW +Y+ P LL++FTKETGI+V Y T++SNE MY
Sbjct: 8   LLAAGALALGMSAAHADDNNTLYFYNWTEYVPPGLLEQFTKETGIKVIYSTYESNETMYA 67

Query: 77  KIKQ-GGTTYDIAVPSDYTIDKMIKENLLNKLDKSKLVGMDNIGKEFLGKSFDPQNDYSL 135
           K+K     YD+ VPS Y +DKM KE ++ K+DKSKL     N+  + L K FDP NDYS+
Sbjct: 68  KLKTYKDGAYDLVVPSTYYVDKMRKEGMLQKIDKSKLTNFSNLDPDMLNKPFDPNNDYSI 127

Query: 136 PYPWGTVGIVYNDQLVD-KAPMHWEDLWRPEYKNSIIMLIDGAREMLGVGLLTTGFGYSVNSK 194
           PY WG   I  N   VD K+   W DLW+PEYK S++L D ARE+  + L   GYS N+
Sbjct: 128 PYIWGATAIGVNGDAVDPKSVTSWADLWKPEYKGSLLLTDDAREVFQMALRKLGYSGNTT 187

Query: 195 NLEQLQAAERKLQQLTPNVKAIVADEMKGYMIQGDAAIGITFSGEASEMLDSNEHLHYIV 254
           + ++++AA  +L++L PNV A  +D      ++G+  +G+ ++G A     +  +
Sbjct: 188 DPKEIEAAYNELKKLMPNVAAFNSDNPANPYMEGEVNLGMIWNGSAFVARQAGTPIDVVW 247

Query: 255 PSEGSNLWFDNLVLPKTMKHEKRAYAFLNFINRPENAAQNAAYIGYATPNKKAKALLPDE 314
```

```
            P EG   W D+L +P   K+++ A   +NF+ RP+ A Q A  IGY TPN  A+ LL  E
Sbjct: 248 PKEGGIPWMDSLAIPANAKNKGALKLINFLLRPDVAKQVAETIGYTPNLAARKLLSPE 307

Query: 315 IKNDPAFYPTDDIIKKLEVYDNLGSRWLGTYNDLYLQFKMYR 356
           + ND   YP  + IK  E  +++G+     IY + Y + K  R
Sbjct: 308 VANDKTLYPDAETIKNGEWQNDVGAA-SSIYEEYYQKLKAGR 348
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 270/357 (75%), Positives = 306/357 (85%)

Query: 14  MRRVYSFLGGIVLVILLILPGLTTYLEKKSSSTPNSDKLVIYNWGDYIDPALLKKFTKETG 73
           MR++YSFL G++ VI+IL  L+  L+KKS S   SDKLVIYNWGDYIDPALLKKFTKETG
Sbjct: 1   MRKLYSFLAGVLGVIVILTSLSFILQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETG 60

Query: 74  IEVQYETFDSNEAMHTKIKQGGTTYDIAVPSDYMIDKMIKNLLVKLDHSKIANWDAIGA 133
           IEVQYETFDSNEAM+TKIKQGGTTYDIAVPSDY IDKMIKNLL KLD SK+   D IG
Sbjct: 61  IEVQYETFDSNEAMYTKIKQGGTTYDIAVPSDYTIDKMIKENLLNKLDKSKLVGMDNIGK 120

Query: 134 RFKNLSFDPKNKYSIPYFWGTVGIVYNDQLVKTPPKHWDDLWRPEFRNKIMLVDSAREVI 193
            F   SFDP+N YS+PYFWGTVGIVYNDQLV    P HW+DLWRPE++N IML+D ARE++
Sbjct: 121 EFLGKSFDPQNEYSLPYFWGTVGIVYNDQLVDKAPMHWEDLWRPEYKNSIMLIDGAREMI 180

Query: 194 GVGLNSLGYGLNTKNISKLKAASKKLDALTPNVKAIVADEMKGYMIQSDAAIGVTPSGEA 253
           GVGL + GY +N+KN+ +L+AA +KL  LTPNVKAIVADENKGYMIQSDAAIG+TPSGEA
Sbjct: 181 GVGLTTFGYSVNSKNLEQLQAARKKLQQLTPNVKAIVADENKGYMIQSDAAIGITPSGEA 240

Query: 254 REMLDGNKHLHYVVPSEGSNLWFDNIVIPKTVKHRKEAYAFINFWMEPKNAAQNAEYIGY 313
            EMLD N+HLHY+VPSEGSNLWFDN+V+PKT+KH KEAYAF+NF+  P+NAAQNA YIGY
Sbjct: 241 SEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYAFLNFINRPENAAQNAAYIGY 300

Query: 314 ATPNLKAKALLPADIKNDKAFYPPDETIDHLEVYNNLGQKWLGIYNDLYLQFKMYRK 370
           ATPN KAKALLP +IKND AFYP D  I  LEVV+NLG +WLGIYNDLYLQFKMYRK
Sbjct: 301 ATPNKKAKALLPDEIKNDPAFYPTDDIIKKLEVYDNLGSRWLGIYNDLYLQFKMYRK 357
```

SEQ ID 8882 (GBS135) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 6; MW 40kDa).

GBS135-His was purified as shown in Figure 201, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1708

A DNA sequence (GBSx1812) was identified in *S.agalactiae* <SEQ ID 5309> which encodes the amino acid sequence <SEQ ID 5310>. This protein is predicted to be spermidine/putrescine ABC transporter, permease protein (potC). Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.05   Transmembrane   17 -  33 (  10 -  37)
   INTEGRAL    Likelihood = -8.65   Transmembrane  236 - 252 ( 232 - 259)
   INTEGRAL    Likelihood = -7.75   Transmembrane  137 - 153 ( 132 - 158)
   INTEGRAL    Likelihood = -7.17   Transmembrane   63 -  79 (  60 -  92)
   INTEGRAL    Likelihood = -6.32   Transmembrane  108 - 124 ( 107 - 136)

----- Final Results -----
               bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8883> which encodes amino acid sequence <SEQ ID 8884> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 2
SRCFLG: 0
McG: Length of UR:   26
     Peak Value of UR:   3.65
     Net Charge of CR: 2
McG: Discrim Score:     16.58
GvH: Signal Score (-7.5): -6.17
     Possible site: 43
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 4 value: -12.05 threshold:  0.0
   INTEGRAL    Likelihood =-12.05   Transmembrane    9 -  25 (   2 -  29)
   INTEGRAL    Likelihood = -7.75   Transmembrane  129 - 145 ( 124 - 150)
   INTEGRAL    Likelihood = -7.17   Transmembrane   55 -  71 (  52 -  84)
   INTEGRAL    Likelihood = -6.32   Transmembrane  100 - 116 (  99 - 128)
   PERIPHERAL  Likelihood =  0.53     174
 modified ALOM score:   2.91
icm1 HYPID: 7  CFP: 0.582

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB91527 GB:AE001165 spermidine/putrescine ABC transporter,
          permease protein (potC) [Borrelia burgdorferi]
 Identities = 97/249 (38%), Positives = 159/249 (62%), Gaps = 3/249 (1%)

Query: 10  KKFANIYLALVFIILYIPIIYLIFYSFNKGGDMNSFTGFTFSHYGELPQDSRLMLILVQT 69
           + F NI+L L+   +Y+PII LI YSFN G     + GF+   Y E+P  S++    +  T
Sbjct: 3   RAFKNIFLFLILSFIYLPIIILIIYSFNSGDSGFIWQGFSLKWYKSIFASSQIKSAIFNT 62

Query: 70  FFLAFLSALLATIIGTFGAIWIVQVRRRH-QTSILSLNNILLVAPDVMIGASFLLVFTVI 128
           +A +S+L + +IG  GA  IY+  +  +T +LS+N I ++ PD++ G S +  ++ I
Sbjct: 63  ILIAIISSLTSVVIGIIGAYAIYKSENKKLKTILLSVNKITIINPDIVTGISLMTFYSAI 122

Query: 129 GLQLGFTSVLLSHVAFSIPIVVLMVLPRLKEMNDDMINASYDLGASTWQMLKEVMLPYLS 188
            +QLGF++LL+SH+ FS P VV++LP+L  + ++I+A+ DLGAS  Q+   ++ P ++
Sbjct: 123 KMQLGFSTMLISHIIFSTPYVVIILPKLYSLFKNIIDAAKDLGASRIQIFFNIIYPEIA 182

Query: 189 SGIISGFFWAFTYSLDDFAVTFFVTGNGFSTLSVEIYSRARRGISLEINALSTIVF--LF 246
             I +G  +AFT S+DDF ++FF TG GF+ LS+ I S  +RGI   IN+S I+F  +
Sbjct: 183 GSIATGALIAFTLSIDDFLISFFTTGQGFNNLSILINSLTKRGIKPVINAISAILFTIL 242

Query: 247 SILLVIGYY 255
           S+L +I  +
Sbjct: 243 SLLFIINKF 251
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5311> which encodes the amino acid sequence <SEQ ID 5312>. Analysis of this protein sequence reveals the following:

```
     Possible site: 49
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.17   Transmembrane    9 -  25 (   4 -  29)
   INTEGRAL    Likelihood = -8.12   Transmembrane  228 - 244 ( 224 - 250)
   INTEGRAL    Likelihood = -7.91   Transmembrane  129 - 145 ( 124 - 150)
   INTEGRAL    Likelihood = -7.06   Transmembrane   62 -  78 (  54 -  87)
   INTEGRAL    Likelihood = -3.93   Transmembrane  100 - 116 (  99 - 118)
```

```
----- Final Results -----
                bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB91527 GB:AE001165 spermidine/putrescine ABC transporter,
          permease protein (potC) [Borrelia burgdorferi]
 Identities = 91/249 (36%), Positives = 154/249 (61%), Gaps = 3/249 (1%)

Query: 2   KKFANLYLASVFVLLYIPIFYLIFYSFNKGGDMNGFTGFTLEHYQIMFEDSRLMTILLQT 61
           + F N++L  +    +Y+PI  LI YSFN G      + GF+L+ Y+ +F  S++ + +  T
Sbjct: 3   RAFKNIFLPLLLSFIYLPIILIIYYSFNSGDSGFIWQGFSLKWYKEIFASSQIKSAIFNT 62

Query: 62  FVLAFSSALLATIIGIPGAIPIHHVRGK-YQNAMLSANNVLMVSPDVMIGASFLILFTSL 120
            ++A  S+L + +IGI GA  I+    K  +  +LS N + +++PD++ G S +  ++++
Sbjct: 63  ILIAIISSLTSVVIGIIGAYAIYKSENKKLKTILLSVNKITIINPDIVTGISLMTFYSAI 122

Query: 121 KPQLGMSSVLLSHIAFSIPIVVLMVLPRLKEMNQDMVNAAYDLGANYPQMLKEVMLPYPT 180
           K QLG S++L+SHI FS P VV+++LP+L  + +++++AA DLGA+ Q+   ++ P
Sbjct: 123 KMQLGFSTMLISHIIFSTPYVVIIILPKLYSLPKNIDAAKDLGASEIQIFFNIIYPEIA 182

Query: 181 PGIIAGYFWAFTYSLDDFAVTFFLTGNSVTTLSVEIYSRARQGISLDINALSTIVFF--F 238
            I  G  +APT S+DDF ++FF TG     LS+ I S  ++GI   INA+S I+FF
Sbjct: 183 GSIATGALIAFTLSIDDFLISFFTTSQGFNNLSILINSLTKRGIKFVINAISAILFFTIL 242

Query: 239 SILLVIGYY 247
           S+L +I  +
Sbjct: 243 SLLFIINKF 251
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 196/258 (75%), Positives = 231/258 (88%)

Query: 9   MKKFANIYLALVFILYIPIYLIFYSFNKGSDMNSFTGFTFSHYGELFQDSRLMLILVQ 68
           MKKFAN+YLA VF++LYIPI YLIFYSFNKGSDMN FTGFT  HY  +F+DSRLM IL+Q
Sbjct: 1   MKKFANLYLASVFVLLYIPIFYLIFYSFNKGSDMNGFTGFTLEHYQTMFEDSRLMTILLQ 60

Query: 69  TFFLAFLSALLATIIGTPGAIWIYQVRRRHQTSILSLNNILLVAPDVMIGASFLLVFTVI 128
           TF LAF SALLATIIG PGAI+I+ VR ++Q ++LS NN+L+V+PDVMIGASFL++FT +
Sbjct: 61  TFVLAFSSALLATIIGIPGAIFIHHVRGKYQNAMLSANNVLMVSPDVMIGASFLILFTSL 120

Query: 129 GLQLGFTSVLLSHVAFSIPIVVLMVLPRLKEMNDDMINASYDLGASTWQMLKEVMLPYLS 188
             QLG +SVLLSH+AFSIPIVVLMVLPRLKEMN DM+NA+YDLGA+ +QMLKEVMLPY +
Sbjct: 121 KPQLGMSSVLLSHIAFSIPIVVLMVLPRLKEMNQDMVNAAYDLGANYPQMLKEVMLPYFT 180

Query: 189 SGIISGFFMAFTYSLDDFAVTFFVTGNGFSTLSVEIYSRARRGISLEINALSTIVFLFSI 248
            GII+G+FMAFTYSLDDFAVTFF+TGN  +TLSVEIYSRAR+GISL+INALSTIVF FSI
Sbjct: 181 PGIIAGYFMAFTYSLDDFAVTFFLTGNSVTTLSVEIYSRARQGISLDINALSTIVFFFSI 240

Query: 249 LLVIGYYYISKEKGEKNA 266
           LLVIGYYY+S++K EK+A
Sbjct: 241 LLVIGYYYMSQDKEEKHA 258
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1709

A DNA sequence (GBSx1813) was identified in *S.agalactiae* <SEQ ID 5313> which encodes the amino acid sequence <SEQ ID 5314>. This protein is predicted to be spermidine/putrescine ABC transporter, permease protein (potB). Analysis of this protein sequence reveals the following:

```
Possible site: 35
```

```
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.55   Transmembrane  250 - 266 ( 244 - 269)
   INTEGRAL    Likelihood = -3.93   Transmembrane  148 - 164 ( 146 - 166)
   INTEGRAL    Likelihood = -3.35   Transmembrane   65 -  81 (  64 -  85)
   INTEGRAL    Likelihood = -1.97   Transmembrane   96 - 112 (  96 - 115)

----- Final Results -----
               bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9853> which encodes amino acid sequence <SEQ ID 9854> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC22990 GB:U32813 spermidine/putrescine ABC transporter,
          permease protein (potB) [Haemophilus influenzae Rd]
 Identities = 90/255 (35%), Positives = 153/255 (59%), Gaps = 11/255 (4%)

Query: 21  AWLFLFVLAPVALIAWNSFFDINSH------FTLANYQTFFSSGTYLKMSPNSVLYAGIV 74
           +WL  FVL P  L+   SF   +G        T+ NY   F+   Y ++ +NS+  +GI
Sbjct: 18  SWLIFVLIPNLLVLAVSPLTRDGSNFYAFPITENYTNLFNP-LYAQVVWNSLSMSGIA 76

Query: 75  SFITLLISYPAAYLLTKL--KHKQLWLMLVILPTWINLLLKAYAFMGIFPGQQGGINAFLT 132
           + I LLI YP A+++K+  K++ L L LV+LP W N L++ Y      G +G +N  L
Sbjct: 77  TIICLLIGYPFAFMMSKIHPKYRPLLLFLVVLPFWTNSLIRIYGMKVFLGVKGILNTMLI 136

Query: 133 FIGI--GPKQILFTDFSFLFVAAYIELPFMLLPIFNALDDIDQNLIYASDDLGANAWQTF 190
             +GI   P +IL T+ + +     Y+ LPFM+LP++++A++ +D  L+ A+ DLGAN +Q F
Sbjct: 137 DMGILSAPIRILNTEIAVIIGLVYLLLPFMILPLYSAIEKLDNRLLEAARDLGANTFQRF 196

Query: 191 QKVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAIEQHFLITQNKGMGST 250
            +VI PL++ G+ AG   V +P++ +F +  L+GG +V+ +G  I+  FLI++N    GS
Sbjct: 197 FRVILPLTMPGIIAGCLLVLLPAMGMFYVADLLCGAKVLLVGNVIKSEFLISRNWPFGSA 256

Query: 251 IGVILILVMVAIMWL 265
           + + L ++M  ++++
Sbjct: 257 VSIGLTVLNALLIFV 271
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5315> which encodes the amino acid sequence <SEQ ID 5316>. Analysis of this protein sequence reveals the following:

```
      Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.38   Transmembrane   19 -  35 (  11 -  40)
   INTEGRAL    Likelihood = -6.79   Transmembrane  250 - 266 ( 245 - 268)
   INTEGRAL    Likelihood = -4.83   Transmembrane   65 -  81 (  63 -  85)
   INTEGRAL    Likelihood = -1.97   Transmembrane   96 - 112 (  96 - 115)
   INTEGRAL    Likelihood = -1.91   Transmembrane  148 - 164 ( 148 - 165)

----- Final Results -----
               bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC22990 GB:U32813 spermidine/putrescine ABC transporter,
          permease protein (potB) [Haemophilus influenzae Rd]
 Identities = 91/262 (34%), Positives = 158/262 (59%), Gaps = 11/262 (4%)

Query: 20  FLWILFPVVAPVTLLPYKSPFDIEGR------VTLANYETPFSSWTYLRMSVNSILYAGI 73
           F W++PFV+ P  L+   SF   +G        +T+ NY   F+   Y ++ NS+  +GI
Sbjct: 17  FSWLIPFVLIPNLLVLAVSPLTRDGSNFYAFPITIENYTNLFNP-LYAQVVWNSLSMSGI 75
```

```
Query: 74  ITLVTLLISYPTALFLTRL--KHKQLWLMLIILPTWVNLLLKAYAPMGIFGQQGGINSFL 131
           T++ LLI YP A  ++++  K++ L L L+LP W N L++ Y      G +G +N+ L
Sbjct: 76  ATIICLLIGYPFAPWMSKIHPKYRPLLLFLVVLPFWTNSLIRIYGMKVFLGVKGILNTNL 135

Query: 132 TFMGI--GPQQILFTDFSFIFVASYISLPFMMLPIFNALDDIDHNVINASRDLGASEFQA 189
             MGI    P +IL T+ + I     Y+ LPFM+LP++++A++ +D+ ++ A+RDLGA+ FQ
Sbjct: 136 IDMGILSAPIRILNTRIAVIIGLVYLLLPFMILPLYSAIEKLDNRLLEAARDLGANTFQR 195

Query: 190 FSKVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGKNRVITLGTPAIEQHFLTTQNWKMGS 249
           F +VI PL++ G+ AG   V +P++ +F +  L+GG +V+ +G  I+  FL ++NW  GS
Sbjct: 196 FFRVILPLTMPGIIAGCLLVLLPAMGMFYVADLLGGAKVLLVGNVIKSEFLISRNWPFGS 255

Query: 250 TIGVVLILTMVAIMWLTKEKSK 271
             + + L + M  ++++     +K
Sbjct: 256 AVSIGLTVLMALLIFVYYRANK 277
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 215/266 (80%), Positives = 239/266 (89%)

Query: 4   RRREMKKTSSLPSIPYMAWLFLFVLAPVALIAWNSPFDINGHPTLANYQTPFSSGTYLKM 63
           RR  MKKTSSLPSIPY  W+  FV+APV L+ + SPFDI G  TLANY+TPFSS TYL+M
Sbjct: 4   RRSVMKKTSSLPSIPYFLWILFFVVAPVTLLFYKSPFDIEGRVTLANYETPFSSWTYLRM 63

Query: 64  SFNSVLVAGIVSPITLLISYPAAYLLTKLKHKQLWLMLVILPTWINLLLKAYAPMGIFGQ 123
           S NS+LVAGI++ +TLLISYP A  LT+LKHKQLWLML+ILPTW+NLLLKAYAPMGIFGQ
Sbjct: 64  SVNSILVAGIITLVTLLISYPTALFLTRLKHKQLWLMLIILPTWVNLLLKAYAPMGIFGQ 123

Query: 124 QGGINAPLTFIGIGPKQILFTDFSFLFVAAYIELPFMLLPIFNALDDIDQNLIYASDDLG 183
           QGGIN+PLTF+GIGP+QILFTDFSF+FVA+YIELPFM+LPIFNALDDID N+I AS DLG
Sbjct: 124 QGGINSFLTFMGIGPQQILFTDFSFIFVASYIELPFMMLPIFNALDDIDHNVINASRDLG 183

Query: 184 ANAWQTFQKVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAIEQHFLITQ 243
           A+ +Q F KVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAIEQHFL TQ
Sbjct: 184 ASEFQAFSKVIFPLSLNGVRAGVQSVFIPSLSLFMLTRLIGGNRVITLGTAIEQHFLTTQ 243

Query: 244 NKGMGSTIGVILILVMVAIMWLTKER 269
           N GMGSTIGV+LIL MVAIMWLTKE+
Sbjct: 244 NWGMGSTIGVVLILTMVAIMWLTKEK 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1710

A DNA sequence (GBSx1814) was identified in *S.agalactiae* <SEQ ID 5317> which encodes the amino acid sequence <SEQ ID 5318>. This protein is predicted to be spermidine/putrescine ABC transporter, ATP-binding protein (potA). Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3031(Affirmative) < Succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < Succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < Succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB91525 GB:AE001165 spermidine/putrescine ABC transporter,
           ATP-binding protein (potA) [Borrelia burgdorferi]
 Identities = 166/345 (48%), Positives = 240/345 (69%), Gaps = 1/345 (0%)

Query: 1   MTNPIIAFKNVSKVFEDSNTVVLKDINFSLEEGKFVTLLGASGSGKSTILNIIAGLLEAS 60
           M N I+  KN+S  ++++    L +IN ++++ +F TLLG SG GK+T++ I+ G L
```

```
Sbjct: 1   MDNCILEIKNLSHYYDNNGNKTLDNINLKIKKNEFITLLGPSGCGKTTLLKILGGFLSQK 60

Query: 61  TGDIYLDGKRINDVPTNKRDVFTVFQNYALFPHMTVFENVAFPLKLKKMDKKEIQKRVQE 120
            G+IY   K I+    NKR++++TVFQNYALFPHM VF+N++F LK+KK  K  I+++V+
Sbjct: 61  NGEIYFFSKEISKTSPNKREINTVFQNYALFPHMNVFDNISFGLRMKKTPKDIIKSKVKT 120

Query: 121 TLKMVRLEGFEKRAIQKLSGGQQRQRVAIARAIINQPKVVLLDEPLSALDLKLRTEMQYEL 180
            +L ++ +  + R I +LSGGQ+QRVAIARA++ +PK++LLDEPLSALDLK+R EMQ EL
Sbjct: 121 SLSLIGMPKVAYRNINELSGGQKQRVAIARAMVMEPKLLLLDEPLSALDLKMRQEMQKEL 180

Query: 181 RELQQRLGITFVFVTHDQEEALAMSDMIFVMNEGEIVQSGTPVDIYDEPINHFVATFIGE 240
            +++Q++LGITF++VTHDQEEAL MSD I VMNEG I+Q GTP +IY+EP   FVA FIGE
Sbjct: 181 KKIQRQLGITFIYVTHDQEEALTMSDRIVVMNEGIILQIGTPEEIYNEPKTKFVADFIGE 240

Query: 241 SNILSGKMIEDYLVEPNGKRFEAVDGGMRNESVQVVIRPEDLQITLPDEGKLQVKVDTQ 300
            SNI  G   ++ +V   G  FE +D G    E+V +VIRPED+++    +G L   + +
Sbjct: 241 SNIFDGTYKKELVVSLLGHEPECLDKGPSAEEAVDLVIRPEDVKLLPKGKGHLSGTITSA 300

Query: 301 LFRGVHYEIIAYDDLGNEWMIHSTRKAIEGEVIGLDFTPEDIHIM 345
            +F+GVHYE+        N W++ STR    GE + +   P+DIH+M
Sbjct: 301 IFQGVHYEMTLEIQKTN-WIVQSTRLTKVGEEVDIFLEPDDIHVM 344
```

There is also homology to SEQ ID 1292

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1711

A DNA sequence (GBSx1815) was identified in *S.agalactiae* <SEQ ID 5319> which encodes the amino acid sequence <SEQ ID 5320>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4990(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06283 GB:AP001515 UDP-N-acetylenolpyruvoylglucosamine
            reductase [Bacillus halodurans]
 Identities = 119/286 (41%), Positives = 166/286 (57%), Gaps = 1/286 (0%)

Query: 13  DIRFDEPLKKYTYTKVGGPADYLAFPRNRLELSRIVKFANSQNIPMMVLGNASNLIVRDG 72
            ++R +E L  +T  K+GGPAD     P +   L  +K        W V+G  SNI+V D
Sbjct: 15  EVRVNESLAHHTTWKIGGPADVFVIPNDIEGLKNTMKLIQENGCKWRVIGRGSNILVSDK 74

Query: 73  GIRGFVIMPDK-LSTVTVNGYVIRAEAGANLIETRIARYHSLTGPEFACGIPGSVGGAV 131
            G+RG I DK L + VNG I  AG +++   +     L G EFA GIPGSVGGAV
Sbjct: 75  GLRGVTIKLDKGLDHLEVNGESITVGAGPFVVKIATVISRQGLAGLEFAAGIPGSVGGAV 134

Query: 132 FMNAGAYGGEIAHILLSAQVLTPQGELKTIRARNMQFGYRHSVIQESGDIVISAKFALKP 191
            FMNAGA+G +I+ IL  A VL P G L+ +     M F YR S++Q++  I + A F+L
Sbjct: 135 FMNAGAHGSDISQILTKAHVLFPDGTLRWITNKEMAFSYKTSLLQKNDGICVKAIFSLTR 194

Query: 192 GDHLMITQEMDRLTYLKELKQPLEYPSCGSVFKRPGHFAGQLISEAHLKGQRIGGVEVS 251
            GD   I +++ +   R  QP  +P+CGSVF+ P  +AGQLI +A LKG +IGG ++S
Sbjct: 195 GDKEDIKKKLQNKDYRRDTQPWNHPTCGSVFRNPLPEYAGQLIEKAGLKGYQIGGAQIS 254

Query: 252 QKHAGFMVNIAEGSAQDYENLIEHVINTVESTSGVHLEPEVRIIGE 297
             HA F+VN  + A D   LI HV +T++    +++E EV +IGE
Sbjct: 255 TMHANFIVNTGDAKAADVLALIHHVKDTIQKQYGMNMETEVELIGE 300
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5321> which encodes the amino acid sequence <SEQ ID 5322>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.4557(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 229/292 (78%), Positives = 267/292 (91%)

Query: 8   ELEGLDIRFDRPLKKYTTKVGGPADYLAPPRNRLESRIVKFANSQNIPWMVLGNASNI 67
           EL G+DIR +EPLK YTYTKVGGPAD+LAPPRN  ELSRIV +AN +N+PW+VLGNASN+
Sbjct: 4   ELHGIDIRENEPLKHYTYTKVGGPADFLAPPRNHYELSRIVAYANKENMPWLVLGNASNL 63

Query: 68  IVRDGGIRGFVIMPDKLSTVTVNGYVIEAEAGANLIETTRIARYHSLTGFEFACGIPGSV 127
           IVRDGGIRGFVIMPDKL+ V +NGY +EAEAGANLIETT+IA++HSLTGFEFACGIPGS+
Sbjct: 64  IVRDGGIRGFVIMPDKLNAVHLNGYTLEAEAGANLIETTKIAKFHSLTGFEFACSIPGSI 123

Query: 128 GGAVFMNAGAYGCEIAHILLSAQVLTPQGELKTIEARNMQFGYRHSVIQESGDIVISAKF 187
           GGAVFMNAGAYGGEI+HI LSA+VLTP GE+KTI AR+M FGYRHS IQE+GDIVISAKF
Sbjct: 124 GGAVFMNAGAYGGEISHIFLSAKVLTPSGKIKTISARDMAFGYRHSAIQETGDIVISAKF 183

Query: 188 ALKPGDHLMITQEMDRLTYLRELKQPLEYPSCGSVFKRPPGHFAGQLISEAHLKGQRIGG 247
           ALKPG++  I+QEM+RL +LR+LKQPLE+PSCGSVFKRPPGHFAGQLI EA+LKG RIGG
Sbjct: 184 ALKPGNYDTISQEMNRLNHLRQLKQPLEFPSCGSVFKRPPGHFAGQLIMEANLKGHRIGG 243

Query: 248 VEVSQKHAGFMVNIAEGSAQDYENLIEHVINTVSSTSGVHLEPEVRIIGESL 299
           VEVS+KH GFM+N+A+G+A+DYE+LI +VI TVS+ SGV LEPEVRIIGE+L
Sbjct: 244 VEVSEKHTGFMINVADGTAKDYEDLIAYVIETVENHSGVRLEPEVRIIGENL 295
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1712

A DNA sequence (GBSx1816) was identified in *S.agalactiae* <SEQ ID 5323> which encodes the amino acid sequence <SEQ ID 5324>. This protein is predicted to be 2-amino-4-hydroxy-6-hydroxymethyldihydropterin pyrophosphokinase/dihyd. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.1122(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB03814 GB:AP001507
           2-amino-4-hydroxy-6-hydroxymethyldihydropteridin e
           pyrophosphokinase [Bacillus halodurans]
 Identities = 64/146 (43%), Positives = 94/146 (63%)

Query: 5   YLSLGSNIGDRETFLKQALPSIDHLQKTKVAQISAIYETAAWGNTNQEDFFNICCQVETD 64
           Y++LGSNIGDR  FL++A+  +    K  V   S+IYET   G T+Q  F N+  +V T
Sbjct: 6   YIALGSNIGDRSRFLESAIQQLAEHDKVTVTCCSSIYETDPVGYTDQSPFLNMVVEVSTS 65
```

```
Query: 65  LAPFELLDYCQRIEKCLKRVRHEHWGPRTIDIDILLFGNQVINQSDLVVPHPYMTKRAFV 124
           L   +LL+  Q+IE+   R RH  WGPRT+D+DILL+  +     E+L++PHP M +RAFV
Sbjct: 66  LPVEQLLEVTQKIERYCGRERHIRWGPRTLDLDILLYDQENREMNLLIPHPRMWERAPV 125

Query: 125 LVPLLEIAPQLSLPNGSKLEDYLEKL 150
           L+PL+E+ P +  P+G  +E  + +L
Sbjct: 126 LIPLMELNPSIVAPSGKTIEQVVREL 151
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5325> which encodes the amino acid sequence <SEQ ID 5326>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0479(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 85/156 (54%), Positives = 111/156 (70%), Gaps = 1/156 (0%)

Query: 1   MTTVYLSLGSNIGDRETFLKQALFSIDHLGKTKVAQISAIYETAAWGNINQDDFFNICCQ 60
            MT VYLSLG+N+GDR  +L++AL ++   L +T++    S+IYET AWG T Q DF N+ CQ
Sbjct: 1   MTIVYLSLGTNMGDRAAYLQKALEALADLPQTRLLAQSSIYETTAWGKTGQADFLNMACQ 60

Query: 61  VETDLAPFELLDYCQEIEKCLKRVRHEHWGPRTIDIDILLFGNQVINQEDLVVPHPYMTK 120
           ++T L   + L   Q IE+ L RVRHE WG RTIDIDILLFG +V + ++L VPHPYMT+
Sbjct: 61  LDTQLTAADFLKETQAIEQSLGRVRHEKWGSRTIDIDILLFGEEVVDTKELKVPHPYMTE 120

Query: 121 RAFVLVPLLEIAPQLSLPNGSK-LEDYLEKLNLGEV 155
           RAFVL+PLLE+ P L LP    K L DYL  L+  ++
Sbjct: 121 RAFVLIPLLELQPDLKLPPNHKFLRDYLAALDQSDI 156
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1713

A DNA sequence (GBSx1817) was identified in *S.agalactiae* <SEQ ID 5327> which encodes the amino acid sequence <SEQ ID 5328>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2826(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5329> which encodes the amino acid sequence <SEQ ID 5330>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3547(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 75/119 (63%), Positives = 92/119 (77%)

Query: 1   MDKIYLNKCRFYGYHGAPSEEQTLGQVFQVDAVLSLDLAKASQTDDLIDTVHYGEVFDCI 60
           MDKI L  CRFYGYHGAP EEQTLGQ+F VD  LS+DL  AS +D L DTVHYG VFD +
Sbjct: 1   MDKIVLEGCRFYGYHGAFKEEQTLGQIFLVDLELSVDLQAASLSDQLIDTVHYGMVFDSV 60

Query: 61  KNHVRNEQYQLIEKLAGVIVEDIFLQFHFVQAITLKITKDNPPINGHYESVGIELERRR 119
           +  VE E++ LIE+LAG I E +F +F P++AI + I K+NPPI GHY++VGIELER+R
Sbjct: 61  RQLVEGEKPILIERLAGAICEQLFNREPPPIKAIKVAIKKDKNPPIAGHYKAVSIELERQR 119
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1714

A DNA sequence (GBSx1818) was identified in *S.agalactiae* <SEQ ID 5331> which encodes the amino acid sequence <SEQ ID 5332>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5333> which encodes the amino acid sequence <SEQ ID 5334>. Analysis of this protein sequence reveals the following:

```
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 181/267 (67%), Positives = 224/267 (83%), Gaps = 1/267 (0%)

Query: 1   MKIGQYDITGKACIMGILNVTPDSFSDGGSYTTIDSALNQVGEMLEQGVAIVDIGGESTR 60
           MKIG++ I G A IMGILNVTPDSFSDGGSYTT+  AL+ V +M+  G  I+D+GGESTR
Sbjct: 1   MKIGKFVIEGNAAIMGILNVTPDSFSDGGSYTTVQKALDHVEQMIADGANIIDVGGESTR 60

Query: 61  PGAVFVTAEEEIKRVVPMIKAIRSVVPDLLLSIDTYKTEVAQAALDAGVHILNDVWSGLY 120
           PG  FV+A +EI RVVP+IKAI+S Y D+L+SIDTYKTE A+AAL+AG  ILNDVW+GLY
Sbjct: 61  PGCQFVSATDEIDRVVPVIKAIKENY-DILISIDTYKTETARAALEAGADILNDVWAGLY 119

Query: 121 DGKMLSLAAERNVPIILMHNQEEAVYQDIKKEVCEFLLERAERALEAGVSKDNIWIDPGF 180
           DG+M +LAAE + PIILMHNQ+E VYQ++ ++VC+FL  RA+ AL+AGV K+NIW+DPGF
Sbjct: 120 DGQMFALAAEYDAPIILMHNQDEEVYQEVTQDVCDFLGNKAQAALDAGVPKNNIWVDPGF 179

Query: 181 GFAKTEEQNLELLKGLEQVCDLGYPVLPGISRKRTVNYLLGGNREVTERDMGTAALSAWA 240
           GFAK+ +QN ELLKGL++VC LGYPVLPGISRKR V+ LLGGN +  ERD  TAALSA+A
Sbjct: 180 GFAKSVQQNTELLKGLDRVCQLGYPVLPGISRKRVVDALLGGNTKAKERDGATAALSAYA 239

Query: 241 IAKGCQIVRVHNVEVNKDIVTVISQLV 267
           + KGCQIVRVH+V+ N+DIV V+SQL+
Sbjct: 240 LGKGCQIVRVHDVKANQDIVAVLSQLM 266
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1715

A DNA sequence (GBSx1819) was identified in *S.agalactiae* <SEQ ID 5335> which encodes the amino acid sequence <SEQ ID 5336>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2429(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5337> which encodes the amino acid sequence <SEQ ID 5338>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1590(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 151/184 (82%), Positives = 166/184 (90%)

Query: 3   NQEKMEKAIYQFLEALGENPNREGLKDTPKRVAKMYIEMFSGLNQDPKEQFTAVFSENHE 62
           N+EK E AIYQFLEA+GENPNREGL DTPKRVAKMY EMF GL +DPKE+FTAVF E HE
Sbjct: 16  NKEKAEAAIYQFLEAIGENPNREGLLDTPKRVAKMYAEMFLGLGKDPKEEFTAVFKEQHE 75

Query: 63  EVVIVKDIPFYSMCEHHLVPFYGKAHIAYLPNDGRVTGLSKLARAVEVASKRPQLQERLT 122
           +VVIVKDI FYS+CEHHLVPFYGKAHIAYLP+DGRVTGLSKLARAVEVASKRPQLQERLT
Sbjct: 76  DVVIVKDISFYSICEHHLVPFYGKAHIAYLPSDGRVTGLSKLARAVEVASKRPQLQERLT 135

Query: 123 AQVAQALEDALAPKGIFVMIEAEHMCMTMRGIKKPGSKTITTVARGLYKDDRVERQEILS 182
           +Q+A AL +AL PKG  VM+EAEHMCMTMRGIKKPGSKTITT ARGLYK+ R ERQE++S
Sbjct: 136 SQIADALVEALNPKGTLVMVEAEHMCMTMRGIKKPGSKTITTTARGLYKESRAERQEVIS 195

Query: 183 LIQK 186
           L+ K
Sbjct: 196 LMTK 199
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1716

A DNA sequence (GBSx1820) was identified in *S.agalactiae* <SEQ ID 5339> which encodes the amino acid sequence <SEQ ID 5340>. This protein is predicted to be folylpolyglutamate synthase (folC). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2836(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9855> which encodes amino acid sequence <SEQ ID 9856> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14768 GB:Z99118 folyl-polyglutamate synthetase [Bacillus subtilis]
 Identities = 154/426 (36%), Positives = 245/426 (57%), Gaps = 17/426 (3%)

Query: 3   YQRALEWIHSKLAPGIKPGLERMRWMLEQLGNPQNNLSAIHVVGTNGKGSTTSYLQHIFT 62
            YQ+A  WIH +L PG+KPGL RM+ ++ +LG+P+   + A HV GTNGKGST ++++ +
Sbjct: 5   YQDARSWIHGRLKPGVKPGLGRMKQLMARLGHPEKKIRAFHVAGTNGKGSTVAPIRSMLQ 64

Query: 63  NSGYQVGTFTSPYIVDFRERISIDGQMIPESDFIKLVETVRPVVERLHLERTNLREPATEPE 122
             +GY VGTFTSPYI+ F ERIS++G  I + ++  LV  ++P VE L  +T    TEPE
Sbjct: 65  EAGYTVGTFTSPYIITFNERISVNGIPISDEEWTALVNQMKPIVEALD-QTEYGQPTEPE 123

Query: 123 VITVLMFYYFGNSCPVDIVIIEAGMGGYYDSTNMFKALAVTCPSIGLDHQEVLGRTYVDI 182
            ++T   F YF     VD VI E G+GG +DSTN+ + L     SIG DH  +LG T  +I
Sbjct: 124 IMTACAFLYFAEFHKVDFVIFETGLGGRFDSTNVVEPLLTVITSIGHDHMNILGNTIEEI 183

Query: 183 AEQKVGVLKKGVPFVYANDRQDVEEVFQIKAKETHSQTYRLHNDFYIKEEE-----NYFN 237
            A +K G++K+G+P V A  + +  +V + +A+   +    LH+  I  EE       F+
Sbjct: 184 AGEKAGIIKEGIPIVTAVTQPEALQVIRHEAERHAAPQSLHDACVIPNEEALPAGEQFS 243

Query: 238 YIGPQANIDHIQLQMPGHHQVSNASIAI-TTSLLLRDKYPKLTLQTIKDGLEMTKWVGRT 296
            +   +  + I+  + G HQ  NA++++I    L ++    ++ + ++ GL   W GR
Sbjct: 244 FKTESKCYEDIRTSLIGTHQRQNAALSILAAEWLNKENIAHISDEALRSGLVKAAWPGRL 303

Query: 297 ELI--PPNVMIDGAHNNESVDALVQVIK-KYQQKNVHILFAAINTKPIESMLESLSSIA- 352
            EL+   P V +DGAHN E V+ L + +K ++    + ++F+A+  KP ++M++ L +IA
Sbjct: 304 ELVQEHPFVYLDGAHNEEGVEKLAETMKQRFANSRISVVFSALKDKPYQNMIKRLSTIAH 363

Query: 353 PVSVTSPDYPK-SINLDKYPKAYTRVSDWKKWLHDI-----NLTSDKDPFVITGSLYFIS 406
             +   SFD+P+ S+  D Y +      W +   D+    +     + +ITGSLYFIS
Sbjct: 364 AIHFASPDPFRASLAKDLYDASEISNKSWSEDPDDVIKPIESKKGSNEIVLITGSLYFIS 423

Query: 407 QVRQEL 412
             +R+ L
Sbjct: 424 DIRKRL 429
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5341> which encodes the amino acid sequence <SEQ ID 5342>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.28   Transmembrane    12 -  28 (  12 -  28)

----- Final Results -----
                bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 230/411 (55%), Positives = 295/411 (70%), Gaps = 1/411 (0%)

Query: 1   MTYQEALEWIHSKLAPGIKPGLERMRWMLEQLGNPQNNLSAIHVVGTNGKGSTTSYLQHI 60
            MTY+E LEWIH  L PGIKPGL+RM W+L QLGNPQ N+  +H+VGTNGKGST ++LQHI
Sbjct: 34  MTYEETLEWIHDHLVPGIKPGLKRMLWVLQQLGNPQKNVKGVHIVGTNGKGSTVNHLQHI 93

Query: 61  FTNSGYQVGTFTSPYIVDFRERISIDGQMIPRSDFIKLVETVRPVVERLHLERTNLREPATE 120
            FT +GY+VGTFTSPYI+DF+ERISI+G+MI E D +     +RP+ ERL  ET+    TE
Sbjct: 94  FTTAGYEVGTFTSPYIMDFKERISINGRMISEKDLVIAANRIRPLTERLVQETDPGEVTE 153

Query: 121 PEVITVLMFYYFGNSCPVDIVIIEAGMGGYYDSTNMFKALAVTCPSIGLDHQEVLGRTYV 180
            PEVIT++MF YFG+  PVDI IIEAG+GG YDSTN+F+A+ V CPSIGLDHQ +LG TY
Sbjct: 154 PEVITLIMFLYFGDMHPVDIAIIEAGLGGLYDSTNVFQAMVVVCPSIGLDHQAILGETYA 213

Query: 181 DIAEQKVGVLKKGVPFVYANDRQDVEEVFQIKAKETHSQTYRLHNDFYIKEEENYFNYIG 240
            +IA QK GVL+ G   V+A +     EVF  KA++  +  +   F + E  + + +
```

```
Sbjct: 214 NIAAQKAGVLEGGETLVFAVENPSARELVFLTKAEQVGASIWEWQEQPQMAENASGYRFTS 273

Query: 241 PQANIDHIQLQMPCHHQVSNASIAITTSLLLRDKYPKLTLQTTKDGLEMTKWVGRTELIF 300
           P   I  I + MPCHHQVSNA++AI T L L+D+YP+LT     I++GL  + W+GRTEL+
Sbjct: 274 PLGVISDIHIAMPCHHQVSNAALAIMTCLTLQDRYPRLTPDHIRSGLANSLWLGRTELLA 333

Query: 301 PNVMIDGAHNNESVDALVQVIK-KYQQKNVHILFAAINTKPIESSMLESLSSIAPVSVTSF 359
           PN+MIDGAHNNESV ALV V+K  Y  K +HILF AI+TKPI  ML +L  I  + VTSF
Sbjct: 334 PNLMIDGAHNNESVAALVAVLKNNYNDKKLHILFGAIDTKPIADMLVALEQIGDLQVTSF 393

Query: 360 DYPYKSINLDKYPKAYTRVQDWKKWLHDINLTSDKDFYVITGSLYFISQVRQ 410
           YP +  L+KYP+ + RV+D+K +L          DF+VITGSLYFIS++RQ
Sbjct: 394 HYPNAYPLEKYPERPGRVADFKDFLALRKHAKADDFFVITGSLYFISEIRQ 444
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1717

A DNA sequence (GBSx1821) was identified in *S.agalactiae* <SEQ ID 5343> which encodes the amino acid sequence <SEQ ID 5344>. This protein is predicted to be rarD. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL     Likelihood =-12.31   Transmembrane  130 - 146 ( 125 - 151)
    INTEGRAL     Likelihood =-10.24   Transmembrane  269 - 285 ( 262 - 291)
    INTEGRAL     Likelihood = -7.75   Transmembrane  212 - 228 ( 207 - 233)
    INTEGRAL     Likelihood = -5.52   Transmembrane   80 -  96 (  75 -  99)
    INTEGRAL     Likelihood = -4.14   Transmembrane  106 - 122 ( 104 - 125)
    INTEGRAL     Likelihood = -3.50   Transmembrane  182 - 198 ( 180 - 204)
    INTEGRAL     Likelihood = -2.44   Transmembrane   40 -  56 (  39 -  57)
    INTEGRAL     Likelihood = -0.96   Transmembrane  153 - 169 ( 152 - 169)
    INTEGRAL     Likelihood = -0.32   Transmembrane  251 - 267 ( 250 - 267)

----- Final Results -----
               bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07585 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 109/288 (37%), Positives = 185/288 (63%), Gaps = 6/288 (2%)

Query: 7   GIILGLSAYVLWGLLSLYWKLLSGIEAYSTPAYRIIFTVLTMLIYMLVSGRKTVYLKDLK 66
           G+I  +SAY++WG L LYWKL+  + A    A+RI++++  M+I + V  +    ++++
Sbjct: 8   GVIAAISAYLIWGPLPLYWKLVDEVPASEMLAHRIVWSLGFMVILLAVMKKNRQVMREIL 67

Query: 67  GLVNNKKSPWTMFVASILLSINWLVYIFAVTHGHATEASLGYYMMPIISILLSVLVLREH 126
             + NKK+ + + VA+LLIS+NW ++I+AV+     EASLGYY+ P+I+LLL+++ LRE
Sbjct: 68  DTLANKKTAPGITVAALLISMNWFIFIYAVSSDKVIEASLGYYINPLINVLLAIVFLRES 127

Query: 127 LARVVSLALLIAIMGVGIVVYQTGHFPLISLTLALSFGFVGLLKKSISLSSDFSMLVESS 186
           L++    + L+A  GV  +    G FP ++  LA+SFG YGL+KK +SLS+  S+ +E+
Sbjct: 128 LSKWEVASFLLAAAGVLNITLHYGSFPWVAFALAISFGVYGLIKKVVSLSAWASLTIETL 187

Query: 187 FIAPFALIYIVPF-----AKDFLTDYNILQLVLLSLSGIITAVPLLLFAEAIKRAPLNTI 241
            + PFAL+++++      A  F  ++ +    L+  SG  TA+PLLLFA    KR   ++I
Sbjct: 188 IMTPFALLFLLYIPLSGGASAFSLNH-LSTAWLIIASGAATALPLLLFAVGAKRISFSLI 246

Query: 242 GFIQYINPTIQLLLALFIFKETIVSGEVIGPIFIWLAILVFSIGQVHT 289
           GF+QY+ PTI L+L +F+F+K     + + F+ IW  +++F+I +  T
Sbjct: 247 GFLQYLAPTIMLMLGVFLPQEPPSRVQFVSFLLIWTGLIIFTISRSRT 294
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8885> and protein <SEQ ID 8886> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Cremd: 3
McG: Discrim Score:       5.30
GvH: Signal Score (-7.5): -1.64
     Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 9 value: -12.31 threshold:  0.0
   INTEGRAL    Likelihood =-12.31   Transmembrane  130 - 146 ( 125 - 151)
   INTEGRAL    Likelihood =-10.24   Transmembrane  269 - 285 ( 262 - 291)
   INTEGRAL    Likelihood = -7.75   Transmembrane  212 - 228 ( 207 - 233)
   INTEGRAL    Likelihood = -5.52   Transmembrane   80 -  96 (  75 -  99)
   INTEGRAL    Likelihood = -4.14   Transmembrane  106 - 122 ( 104 - 125)
   INTEGRAL    Likelihood = -3.50   Transmembrane  182 - 198 ( 180 - 204)
   INTEGRAL    Likelihood = -2.44   Transmembrane   40 -  56 (  39 -  57)
   INTEGRAL    Likelihood = -0.96   Transmembrane  153 - 169 ( 152 - 169)
   INTEGRAL    Likelihood = -0.32   Transmembrane  251 - 267 ( 250 - 267)
   PERIPHERAL  Likelihood =  7.96     229
 modified ALOM score:   2.96

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02052(319 - 1152 of 1485)
GP|9654601|gb|AAF93371.1||AE004110(13 - 289 of 302) rarD protein (Vibrio cholerae]
%Match = 20.4
%Identity = 37.7  %Similarity = 66.3
Matches = 104  Mismatches = 89  Conservative Sub.s = 79

117       147       177       207       237       267       297       327
KDIVNLW*RNLK**NKSALKMVRMLLICLEQDRR*WFCVRKKKNKQLSQS*VNYV*VDRFKCLILSEKE*ELRKDNLGII
                                                                           ||:
                                                                    MFMTPDQQDAKKGIL
                                                                             10

357       387       417       441       471       501       531       561
LGLSAVVLWGLLSLYWKLLSGIEAYSTFAYRII--FTVLTMLIYMLVSGRKTVVLKDLVGLVNNKKSFWTMFVASILISI
 | :||| :||: :|:| |  : |  :::|::  | :| :|::   |  |      |:|:  : || ::| ::|:
LAISAYTWWGIADIYFKALGAVSALELSHRVVWSFVLLAVLIHLGRRWRSVV------GVVHTPKKFWLLLVTALLVGG
          30        40        50        60              70        80

591       621       651       681       711       741       771       801
NWLVYIFAVIHGHATEASLGYYMMPIISILLSVLVLRERLARVVSLAILIAIMGVGILVVQTGHPFLISLTLALSFGFYG
|||::|:::   |  :||||||: |::::|| :| | | |  ::  :|: :| :|||| :   |  |:::: || ||||||
NWLIFIWSINANHMLDASLGYYINPLLNVLLGMLFLGERLRKLQWFAVAIAAIGVGIQLVVFGSVPIVAIALATSFGFYG
          100       110       120       130       140       150       160

831       861       891       921       942       972      1002      1032
LLKKSISLSSDFSMLVESSFIAPFALIYIVFFAKDFLTDY--NILQL-VLLSLSGIITAVPLLLFAEAIKRALPNIIGFI
||:| | :  :::|: |: | | ||:::: |  :|  | || :|| :|::| :||| |  |  |: :||
LLRKKIQVDAQTVLPLFLTLFMLFPAAAIYLIWLADITPSDMAINTWQLNLLLVCAGVVTTLPLLCFTGAAARLKLSTLGFF
          180       190       200       210       220       230       240

1062      1092      1122      1152      1182      1212      1242      1272
QYINPTIQLLLALFIFKETIVSGKEVIGFIPIWLAILVFSIGQVHTWLKKGK*DDLSRSARMDS**ISFWY*TRFGTYEMD
||| |:: :|||:::: !  | : | | ||| |:::||:
QYIGPSLMFLLAVLVYGEAFTSDKAITFAFIWSALVIPSVDGLKAGHAARRAR
          260       270       280       290       300
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1718

A DNA sequence (GBSx1822) was identified in *S.agalactiae* <SEQ ID 5345> which encodes the amino acid sequence <SEQ ID 5346>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5200(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1719

A DNA sequence (GBSx1823) was identified in *S.agalactiae* <SEQ ID 5347> which encodes the amino acid sequence <SEQ ID 5348>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0881(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC44297 GB:U41735 homoserine kinase homolog [Streptococcus pneumoniae]
 Identities = 188/289 (65%), Positives = 232/289 (80%), Gaps = 1/289 (0%)

Query: 1   MRIIVPATSANIGPGFDSIGVALSKYLIIEVLEESTEWLVEHNLVN-IPKDHTNLLIQTA 59
           M+IIVPATSANIGPGFDS+GVA++KYL IEV EE  EWL+EH +   IP D  NLL+  A
Sbjct: 1   MKIIVPATSANIGPGFDSVGVAVTKYLQIEVSSERDEWLIEHQIGKWIPHDERNLLITTA 60

Query: 60  LHVKSDLAPHRLKMPSDIPLARGLGSSSSVIVAGIELANQLGNLALSQKEKLSIATRLEG 119
           L +  DL P RLKM SD+PLARGLGSSSSVIVAGIELANQLG L LS  EKL++AT++EG
Sbjct: 61  LQIVPDLQPRRLKMTSDVPLARGLGSSSSVIVAGIELANQLGQNLSDHSKLQLATKIEG 120

Query: 120 HPDNVAPAIFGDLVISSIVKNDIKSLEVMFPDSSFIAPIPNYELKTSDSRNVLPQKLSYE 179
           HPDNVAPAI+G+LVI+S V+  + ++   FP+  F+A+IPNYELT DSR+VLP+KLSY+
Sbjct: 121 HPDNVAPAIYGNLVIASSVESQVSAIVADFPECDFLAYIPNYELRTRDSRSVLPKKLSYK 180

Query: 180 DAVASSSVANVMVASLLKGDLVTAGWAIERDLFHERYRQPLVKEPEVIKQISTQNGAYAT 239
           +AVA+SS+ANV VA+LL GD+VTAG AIE DLFHERYRQ LV+EF +IKQ++ +NGAYAT
Sbjct: 181 EAVAASSIANVAVALLAGDMVTAGQAIEGDLFHERYRQDLVREFAMIKQVTKENGAYAT 240

Query: 240 YLSGAGPTVMVLCSKEKEQAIVTELSKLCLGQQIQVLNIERKGVRVEKR 288
           YLSGAGPTVMVL S +K   I  EL K    G++  L ++ +GVRVE +
Sbjct: 241 YLSGAGPTVMVLASHDEMPTIKAELEKQPFKGKLHDLRVDTQGVRVEAK 289
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1720

A DNA sequence (GBSx1824) was identified in *S.agalactiae* <SEQ ID 5349> which encodes the amino acid sequence <SEQ ID 5350>. This protein is predicted to be homoserine dehydrogenase (hom). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9857> which encodes amino acid sequence <SEQ ID 9858> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA65713 GB:X96988 hom [Lactococcus lactis]
 Identities = 221/432 (51%), Positives = 307/432 (70%), Gaps = 11/432 (2%)

Query: 15  MTIKIALLGFGTVAKGIFYLLKENQHKLLSLSGDEDVIDKVLVRDNESRQRFINQGPTYN 74
           M + IA+LGFGTV  G+P LL EN+ KL  +  E+IVI KVL+RDN++ ++  +QGF Y+
Sbjct: 1   MAVNIAILGFGTVGTGLPTLLSENKEKLAKILDEEIVISKVLMRDNKAIEKARSQGFNYD 60

Query: 75  FVTEINTILQDSQIDIVVELMGGIEPAKTYLSQALGFGKHIVTANKDLIALKGKELMDLA 134
           FV  ++ IL DS+I IVVELMG IEPAKTY++QA+  GK++VTANKDL+A+HG EL  LA
Sbjct: 61  FVLNLDDILADSEISIVVELMGRIEPAKTYITQAIEAGKNVVTANKDLLAVHGVELRSLA 120

Query: 135 DARGLALFYEGAVAGGIPILRTLSHSFASDKMTELLGILNGTSNFMLTMMEEGWSYEQA 194
               +AL+YE AVAGGIPILRTL++SF+SDK+T LLGILNGTSNFM+TKM EEGN+Y+++
Sbjct: 121 QKHHVALYYEAAVAGGIPILRTLANSFSSDKITHLLGILNGTSNFMMTKMSEEGWTYDES 180

Query: 195 LKKAQELGYAESDPTNDVEGIDTAYKATILSQFGFGMPIDPDDVNYKGISSIRSEDVEVA 254
           L KAQELGYAESDPTNDV+GID +YK  ILS+F FGM +  DD+   G+ SI+  DVE+A
Sbjct: 181 LAKAQELGYAESDPTNDVDGIDASYKLAILSEFAFGMTLAPDDIAKSGLRSIQKTDVEIA 240

Query: 255 QEMGFAIKLVADLRETPTGISVDVSPTLISQKHPLAAVNHVMNAVFIESIGIGQSLFYGP 314
           Q+ G+ +KL  ++ E  +GI  +VSPT + + HPLA+VN VMNAVFIES GIG S+FYG
Sbjct: 241 QQFGYVLKLTGRINEVDSGIFAEVSPTFLPKSHPLASVNGVMNAVFIESGIGDSVFYGA 300

Query: 315 GAGQNPTATSVLADIIDISRSIRQQIKIKMNTYHCPCRLSMQSDIFNEYYLAISLRNAE 374
           GAGQ PTATSVLADI+ I + ++     K N Y      L+   DI N+YY ++       E
Sbjct: 301 GAGQKPTATSVLADIVRIVKRVKDGTIGKSNEYARSTSLANPHDIENKYYFSV-----E 355

Query: 375 DSDTLGR------YFEQENIGLKNVIEKALGDKQQEIYVLTDEVSQEKITQFIEEFPESG 428
             D+ G+       F  EN+  + V+++    K+  + +++ ++++ +++   ++  +
Sbjct: 356 TPDSTGQLLLLVELPTSENVSPEQVLQQKGNGKRAVVVIISHKINRVQLSAIQDKLNQEK 415

Query: 429 VIQLINVFKVIG 440
           +L+N FKV+G
Sbjct: 416 DFKLLNRFKVLG 427
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1721

A DNA sequence (GBSx1825) was identified in *S.agalactiae* <SEQ ID 5351> which encodes the amino acid sequence <SEQ ID 5352>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4548(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1722

A DNA sequence (GBSx1826) was identified in *S.agalactiae* <SEQ ID 5353> which encodes the amino acid sequence <SEQ ID 5354>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood =-12.79   Transmembrane    20 -  36 (  14 -  41)

----- Final Results -----
            bacterial membrane --- Certainty=0.6116(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15906 GB:Z99123 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 105/272 (38%), Positives = 149/272 (54%), Gaps = 20/272 (7%)

Query: 25  FLLIALIGIFLFPNNRSKQEIKT-----KTNASSHRKIVTSIKKKK-----WIKQKTPVK 74
           FL I L+G  L     + QE         K       K     ++KK+      WIK + P K
Sbjct: 5   FLSIFLLGSCLALAACADQEANAEQPMPKAEQKKPEKKAVQVQKKEDDTSAWIKTEKPAK 64

Query: 75  IPILMYHAVHVMDPSEAASANLIVAPDIFESHIKRLKKBGYYFLAPNEAYRALNENALPE 134
           +PILMYH++        ++  +L V    FE+H+K L   GY  L P EA   L ++  P
Sbjct: 65  LPILMYHSI-------SSGNSLRVPKKEFEAHMKWLHDNGYQTLTPKEASLMLTQDKKPS 117

Query: 135 KKVIWITFDDGNADFYTKAYPILKKYKVKATNNIITGFVQSGRSENLNVQQMLEMKQNGM 194
           +K + ITFDDG  D Y  AYP+LKKY +KAT  +I    +  G + +L  +QM EM Q+G+
Sbjct: 118 EKCVLITFDDGYTDNYQDAYPVLKKYGMKATIFMIGKSI--GHKHLTEEQMKEMAQHGI 175

Query: 195 SFQGHTVTHPNLSLLTPELQTQRNTLSKQFLDQKLSQDTLAIAYPSGRYNPTTLDIASQY 254
           S + HT+ H  L+ LTP+ Q  EM  SK+  D    Q T  I+YP GRYN   TL  A +
Sbjct: 176 SIESHTIDHLELNGLTPQQQQSEMADSKKLFDNMPHQQTTIISYPVGRYNSETLKAAEKT 235

Query: 255 -YKLGLTTNEGVATKDNGLLSLNRIRILPTTS 285
            Y++G+TT  G A++D G+ +L+R+R+ P  S
Sbjct: 236 GYQMGVTTEPGAASRDQGMYALHRVRVSPGMS 267
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5355> which encodes the amino acid sequence <SEQ ID 5356>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
```

```
>>> May be a lipoprotein

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15906 GB:Z99123 similar to hypothetical protein [Bacillus subtilis]
 Identities = 97/240 (40%), Positives = 140/240 (57%), Gaps = 9/240 (3%)

Query: 71  KKTHPDSSKSQKKAHSKLTWTKQSTPVKIPILMVHAIHVMSPEETANANLIVNEDLFDQQ 130
           KK    + + QKK       W K E P K+PILMVH+I       ++  +L V    F+
Sbjct: 37  KKPEKKAVQVQKKEDDTSAWIKTEKPAKLPILMVHSI-------SSGNSLRVPKKEFEAH 89

Query: 131 LQKMKDEGYYFLSPEEVYRALSNNELPAKKVVWLTFDDSMIDFYNVAYPILKKYDAKATN 190
           ++ + D GY  L+P+E    L+ ++ P++K V +TFDD   D Y  AYP+LKKY  KAT
Sbjct: 90  MKWLHDNGYQTLTPKEASLMLTQDKKPSEKCVLITFDDGYTDNYQDAYPVLKKYGMKATI 149

Query: 191 NVITGLTEMGSAANLTLKQMKEMKQVGMSFQDHTVNHPDLEQASPDVQTTEMKDSKDYLD 250
            +I     +G   +LT +QMKEM Q G+S + HT++H +L    +P  Q +EM DSK   D
Sbjct: 150 FMIG--KSIGHKHHLTEEQMKEMAQHGISIESHTIDHLELNGLTPQQQQSEMADSKKLFD 207

Query: 251 KQLNQFTAIAYPSGRYNDTTLQIAARLNYKLGVTTNEGIASAANGLLSLNRIRLLDNMS 310
            +Q T  I+YP GRYN+ TL+ A +  Y++GVTT G AS   G+ +L+R+R+ P MS
Sbjct: 208 NMFEQQTTIISYPVGRYNEETLKAARKTGYQMGVTTEHGAASRDQGMYALHRVRVSPGMS 267
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 153/265 (57%), Positives = 199/265 (74%), Gaps = 4/265 (1%)

Query: 33  IFLFFNNRSKQEIKTK---TNASSHRKIVTSIKKKKWIKQKTPVKIPILMVHAVHVMDPS 89
           I LF + ++ ++  TK   T+  S +     +  K  W KQ+TPVKIPILMVHA+HVM P
Sbjct: 54  ISLFHHKKTAKKSTTKLKKTHFDSSKSQKKAHSKLTWTKQSTPVKIPILMVHAIHVMSPE 113

Query: 90  EAASANLIVAPDIPESHIKRLKKEGYYFLAPNEAYRALNENALPEKKVIWITFDDGNADF 149
           E A+ANLIV PD+F+  +++++K EGYYFL+P E YRAL+ N LP KKV+W+TFDD   DF
Sbjct: 114 ETANANLIVNPDLPDQQLQKMKDEGYYFLSPEEVYRALSNNELPAKKVVWLTFDDSMIDF 173

Query: 150 YTEAYPILKKYKVKATNNIITGFVQEGRESNLNVQQMLEMKQNGMSFQGHTVTHPNLSLL 209
           Y  AYPILKKY  KATNN+ITG  + G  +NL ++QM EMKQ GMSFQ HTV HP+L
Sbjct: 174 YNVAYPILKKYDAKATNNVITGLTEMGSAANLTLKQMKEMKQVGMSFQDHTVNHPDLEQA 233

Query: 210 TPELQTQEMTLSKQFLDQKLSQDTLAIAYPSGRYNPTTLDIASQY-YKLGLTTNEGVATK 268
           +P++QT EM  SK +LD++L+Q+T+AIAYPSGRYN TTL IA++  YKLG+TTNEG+A+
Sbjct: 234 SPDVQTTEMKDSKDYLDKQLNQNTIAIAYPSGRYNDTTLQIAARLNYKLGVTTNEGIASA 293

Query: 269 DNGLLSLNRIRILPTTSDDDLIKTI 293
            NGLLSLNRIRILP  S ++L++T+
Sbjct: 294 ANGLLSLNRIRILPNMSPENLLQTM 318
```

SEQ ID 5354 (GBS287d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 3 & 4; MW 57kDa) and in Figure 185 (lane 2; MW 57kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 6; MW 32kDa) and in Figure 181 (lane 5; MW 32kDa).

Purified GBS287d-GST is shown in Figure 243, lanes 10-11; purified GBS287d-His is shown in Figure 234, lanes 7-8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1723

A DNA sequence (GBSx1828) was identified in *S.agalactiae* <SEQ ID 5357> which encodes the amino acid sequence <SEQ ID 5358>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1724

A DNA sequence (GBSx1829) was identified in *S.agalactiae* <SEQ ID 5359> which encodes the amino acid sequence <SEQ ID 5360>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3352(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1725

A DNA sequence (GBSx1830) was identified in *S.agalactiae* <SEQ ID 5361> which encodes the amino acid sequence <SEQ ID 5362>. This protein is predicted to be glycine betaine transporter BetL (opuD). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.68   Transmembrane  439 - 455 ( 435 - 491)
   INTEGRAL    Likelihood =-12.10   Transmembrane  256 - 272 ( 249 - 281)
   INTEGRAL    Likelihood =-11.30   Transmembrane  464 - 480 ( 456 - 491)
   INTEGRAL    Likelihood =-10.83   Transmembrane   49 -  65 (  44 -  74)
   INTEGRAL    Likelihood =-10.40   Transmembrane   11 -  27 (   5 -  34)
   INTEGRAL    Likelihood = -9.98   Transmembrane  396 - 412 ( 390 - 419)
   INTEGRAL    Likelihood = -9.29   Transmembrane  224 - 240 ( 220 - 247)
   INTEGRAL    Likelihood = -7.11   Transmembrane  347 - 363 ( 341 - 366)
   INTEGRAL    Likelihood = -2.87   Transmembrane  143 - 159 ( 143 - 159)
   INTEGRAL    Likelihood = -2.60   Transmembrane  192 - 208 ( 191 - 208)
   INTEGRAL    Likelihood = -1.44   Transmembrane   86 - 102 (  86 - 105)

----- Final Results -----
                bacterial membrane --- Certainty=0.6074(Affirmative) < succ>
```

```
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD30266 GB:AF102174 glycine betaine transporter BetL [Listeria
      monocytogenes]
 Identities = 277/503 (55%), Positives = 365/503 (72%), Gaps = 1/503 (0%)

Query: 4   KHITPVFWGSLIVSLILVLLGIIVPRGFQSWTQILREQVSTNFGWLYLLLVTSILALCVF 63
            K +T VF GS  + L+ VL G  +P  F+++T  +++ +++NFGW YL++V  I+  C+F
Sbjct: 2   KKLTNVFWGSGFLVLLAVLFGAFLPSQFETFTNHIQKFLTSNFGWYYLIVVAIIIFCLF 61

Query: 64  FIMSPLGQIRLGQPHSRPEYSTVSWIAMMPSAGMGIGLVFYGAAEPLSHFAISTPGAPKE 123
             ++SP+G IRLG+P    F YS  SW AM+PGAGMGIGLVF+GAAEPLSH+A+  PG
Sbjct: 62  LVLSPIGSIRLGKPGSEPGYSNKSWFAMLFSAGMGIGLVFWGAAEPLSHYAVQAPGGEVG 121

Query: 124 SQTALADAFRFTFPHWGIHAWAVVALVALALAYPGFRKQEKYLLSVTLKPLFGDKTDGWL 183
             +Q A+ DA R++PFHWGI AW++YA+VALALAYF FRK     L+S TL P+ G    G +
Sbjct: 122 TQAAMKDALRYSFPHWGISAWSIYAIVALALAYFKFRKNAPGLISATLYPILGKHAKGPI 181

Query: 184 GKIVDITTVVATVIGVATTLGPGAAQINGGLSFLLGVPNNAFVQIVIILITTALFVMSAL 243
             G+++DI  V ATVIGVATTLG GA QINGGL++L GVPNN  VQ  II+I T LF++SA+
Sbjct: 182 GQLIDIIAVFATVIGVATTLGLGAQQINGGLTYLFGVPNNPTVQFTIIVITLLFMLSAM 241

Query: 244 SGLGKGVKILSNLNLLIAVALLALVIVLGPTVRIFDTLTESLGSYLQNFPGMSFRAAAFD 303
             SGL KG+++LSNN+N+ +A  LL L ++LGPT+ I +  T S G YLQN    MSF+ A
Sbjct: 242 SGLDKGIQLLSNVNIYVAGVLLVLTLILGPTLFIMNNFTNSFGDYLQNIIQMSFQTAFDA 301

Query: 304 NTKRSWIDNWTIFYWAWWISWSPFVGVFIARISKGRSIREFLIVVLLIPTLLSFVWFAAF 363
               R WID+WTIFYWAWW+SWSPFVG+FIARIS+GR+IR+FL  V+++P L+S  WFA F
Sbjct: 302 PDARKWIDSWTIFYWAWWLSWSPFVGIFIARISRGRTIRQFLLGVIVLPALVSVFWFAVF 361

Query: 364 GTLSTQVQQLG-TNLTKFATEEVLFATFNHYTLGWLLSIIAIILIPSFFITSADSATYVL 422
             G  +  V+Q G + L+  ATE+VLF  FN +  G +LSI+A+ILI  FFITSADSAT+VL
Sbjct: 362 GGSAIFVEQHGNSGLSSLATEQVLFGVFNEFPGGMMLSIVAMILIAVFFITSADSATFVL 421

Query: 423 AMLTEDGNLNPKNRTKVIWGLVLAVIAIVLLLSGGLLALQNVLIIVALPFSFVMILMMLA 482
              M T  G+LNP N  KV WGL+ A IA VLL +GGL ALQN  II A PFS V+ILM+++
Sbjct: 422 GMQTTGGSLNPPNSVKVIWGLLQAGIASVLLYAGGLTALQNASIIAAFPFSIVIILMIVS 481

Query: 483 LLVELPHEKKEMGLSISPDRYPR 505
              L V L  E++++GL + P +  R
Sbjct: 482 LFVSLTREQEKLGLYVRPKKSQR 504
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8887> and protein <SEQ ID 8888> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Cremd: 4
McG: Discrim Score:     15.28
GvH: Signal Score (-7.5): -4.24
     Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 11 value: -12.68 threshold:  0.0
   INTEGRAL    Likelihood =-12.68   Transmembrane  439 - 455 ( 435 - 491)
   INTEGRAL    Likelihood =-12.10   Transmembrane  256 - 272 ( 249 - 281)
   INTEGRAL    Likelihood =-11.30   Transmembrane  464 - 480 ( 456 - 491)
   INTEGRAL    Likelihood =-10.83   Transmembrane   49 -  65 (  44 -  74)
   INTEGRAL    Likelihood =-10.40   Transmembrane   11 -  27 (   5 -  34)
   INTEGRAL    Likelihood = -5.98   Transmembrane  396 - 412 ( 390 - 419)
   INTEGRAL    Likelihood = -9.29   Transmembrane  224 - 240 ( 220 - 247)
   INTEGRAL    Likelihood = -7.11   Transmembrane  347 - 363 ( 341 - 366)
   INTEGRAL    Likelihood = -2.87   Transmembrane  143 - 159 ( 143 - 159)
   INTEGRAL    Likelihood = -2.60   Transmembrane  192 - 208 ( 191 - 208)
   INTEGRAL    Likelihood = -1.44   Transmembrane   86 - 102 (  86 - 105)
```

```
  PERIPHERAL  Likelihood =  3.50      319
 modified ALOM score:   3.04

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.6074(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORFG2057(310 - 1821 of 2145)
GP|4835822|gb|AAD30266.1|AF102174_1|AF102174(2 - 506 of 507) glycine betaine transporter
BetL {Listeria monocytogenes} PIR|T48645|T48645 glycine betaine transport protein betL
[validated] - Listeria monocytogenes
%Match = 38.7
%Identity = 54.9  %Similarity = 74.7
Matches = 277  Mismatches = 127  Conservative Sub.s = 100

54          84          114         144         174         204         234         264
IQGGHHYRNYRLEVLKIQDMVVS*ANLDLMPLSTNIWYLHQIVINH*VKHKNQIMLPGSFLLRRQQEVLIQVVKMRGVFI

294         324         354         384         414         444         474         504
KVCYTILV*EEILSKKHITPVTGSLIVSLLIVLLGIIVPERGFQSWTQLIREQVSTNFGWLYLLLVTSILALCVFFIMSP
                  MKKLTNVSWGSGFLVLLAVLFGAFLPEQFETFTNHIQKFLTSNFGWYYLIVVAIIIIFCLFLVLSP
                            10          20          30          40          50          60

534         564         594         624         654         684         714         744
LGQIRLGQPHSRPEYSTVSWIAMMFSAGMGIGLVFYGAAEPLSHFAISTPGAPKESQTALADAPKFTFFHWGIHAWAVYA
IGSIRLGKPGEEPGYSNKSWFAMLFSAGMGIGLVFWGAAEPLSHFAVQAPGSEVGTQAAMKDALRYSFFHWGISAWSIYA
                            80          90          100         110         120         130         140

774         804         834         864         894         924         954         984
LVALALAYFGPRKQEKYLLSVTLKPLPGEKTDGWLGKIVDITTVVATVIGVATTLGPGAAQINGSLSFLLGVPNNAFVQI
IVALALAYFKFRKNAPGLISATLYPILGKHAKGPIGQLIDIIAVFATVIGVATTLGLGAQQINGSLTYLFGVPNNPTVQF
                            160         170         180         190         200         210         220

1014        1044        1074        1104        1134        1164        1194        1224
VIILITTALFVMSALSGLGKGVKILSNLNLIIAVALLALVIVLGPTVRIPDTLTSSLGSYLQNPFGMSFRAAAFDNTKRS
TIIVIVTILFMLSAMSGLDKGIQLLSNVNIYVAGVLLVLTLILGPTLFIMNNFTNSPGDYLQNIIQMSFQTAPDAPDARK
                            240         250         260         270         280         290         300

1254        1284        1314        1344        1374        1404        1431        1461
WIDNWTIFYVNAWWISWSPFVGVPIARISKGRSIREPLIVVLLIPTLLSFVWFAAPGTLSTQVQQLGTN-LTKPATEEVLF
WIDSWTIFYVAWWLSWSPFVGIPIARISKGRTIRQFLLLGVIVLPALVSVFWFAVFGKSAIPVEQHGNSGLSSSLAVTEQVLF
                            320         330         340         350         360         370         380

1491        1521        1551        1581        1611        1641        1671        1701
ATFNHYTLGWLLSIIAILLFSFFITSADSATYVLAMLTEDGNLNPKNRTKVIWGLVLAVIAIVLLLSGGLLALQNVLII
GVFNEPFGGMMLSIVAMILIAVFFITSADSATFVLGMQTTGGSLNPPNSVKVIWGLLQAGIASVLLVYAGGLTALQNASII
                            400         410         420         430         440         450         460

1731        1761        1791        1821        1851        1881        1911        1941
VALPFSFVMILMMLALLVELPHEKKEMGLSISPDRYPRKNEPFKSYEE*KEARRLLFIG*SS*SDHER**LVRYEFD*EK
AAFPPSIVIILMIVSLFVSLTR*QEKLGLYVRPKKSQRSCL
                            480         490         500
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1726

A DNA sequence (GBSx1831) was identified in *S.agalactiae* <SEQ ID 5363> which encodes the amino acid sequence <SEQ ID 5364>. This protein is predicted to be succinic semialdehyde dehydrogenase (gabD-1). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2733(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9859> which encodes amino acid sequence <SEQ ID 9860> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD19405 GB:AF102543 succinic semialdehyde dehydrogenase
            [Zymomonas mobilis]
 Identities = 229/455 (50%), Positives = 305/455 (66%), Gaps = 5/455 (1%)

Query: 10  MAYKTIYPTNEVLHEFDNISDSLEQSLDIAHALYKTWREDNVEERQNQLHKVADLLR 69
           MAY+++ P T E + ++ + SD  ++ S+D A  ++K    +  + ER   LHK A++ R
Sbjct: 1   MAYESVNPATGEIVKKYPDFSDKQVKDSVDRAATVFKNDWSQRTIAERSKVLHKAAEIFR 60

Query: 70  KDRDKYAEVMTKDMGKLFTEAQGEVDLCADIADYYADNGQKFLKPVPLESPNGEAYYLKQ 129
             D DKYA+++T DMGK   EA+GEV+L ADI DYYA NG+KFL P  +E   G A
Sbjct: 61  SDVDKYAKLLTIDMGKKIAEARGEVNLSADILDYYAKNGEKFLAPQKVEEKPG-AVVKAF 119

Query: 130 AVGVLLAVEPWNFPFYQIMRVFAPNFIVGNTMLLKHASICPASAQAFEDLVREAGAPEGA 189
            +G+LLA+EPWNFP+YQ+ R+  P  I GN +L+KH+S  P SA AFE ++ EAGAP+G
Sbjct: 120 PLGLLLAIEPWNFPYYQLARIAGPVLIAGNALLVKHSSSVPQSAHAFEAVLEEAGAPKGI 179

Query: 190 FKNIFASYDQVSNLISDPRVAGVCLTGSERGGASIAABAGKNLKKSSMELGGNDAFLILD 249
           + N+ AS DQVS +I DPRV GV +TGS   GA +AA+AGK  KKS MELGG+DAF++LD
Sbjct: 180 YTNLDASPDQVSQIIEDPRVRGVTVTGSASVGARLAAKAGKWKKSVMELGGSDAFIVLD 239

Query: 250 DADFD--LLSKTIPFARLYNAGQVCTSSKRFIVMADKYDE-FVNMVVETFKSAKWGDPMD 306
             D D  L+ K  +  RL+NAGQV  ++KRFI++  K  E F    + F++ K GDPMD
Sbjct: 240 GVDIDDKLIDKAAY-GRLFNAGQVFCAAKRFIIVGQKRAELFTEKLKQRFSALKIGDPMD 298

Query: 307 SETTLAPLSSAGAKDDVLKQIKLAVDHGAEVVFGNDTIDHPGNFVMPTVLTNITKANPIY 366
             T L PLSS GA+D V+KQ++ AV +GA++V G   I+  G F+   +LT+I + NP Y
Sbjct: 299 ESTDLGPLSSVGARDQVVKQVEKAVQNGAKLVCGGKAIEGKGAFMEAGILTDIKRENPAY 358

Query: 367 NQEIFGPVASIYKVDTEEEAIALANDSSYGLGSTVFSSDPEHAKKVAAQIETGMTFINSG 426
             +E PGP+A IY V  E EAI LANDS YGLG  VF+ D E  +KVA QIETGM VAINKP 418
Sbjct: 359 PEEPFGPIAQIYAVKDKAKAIELANDSPYGLGGAVFAPDVEQGRKVAKQIETGMVAINKP 418

Query: 427 WTSLPELPFGGIKNSGYGRELSQLGFDAFVNEHLV 461
             + PELPFGG+K+SGYGRELS  G   F+N  L+
Sbjct: 419 LWTAPELPFGGVKHSGYGRELSHFGIQEFINWKLI 453
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5365> which encodes the amino acid sequence <SEQ ID 5366>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2887(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 335/457 (73%), Positives = 397/457 (86%)

Query: 9   IMAYKTIYPYTNEVLHEPDNISDSDLEQSLDIAHALYKTWREDNVSERQNQLHKVADLL 68
           +MAY+TIYPYTNEVLH PDN++D  L  L+ AH LYK WREDN+ ERR+ QLH+VA++L
Sbjct: 1   VMAYQTIYPYTNEVLHTPDNMTDQGLADVLERAHLLYKKWREDHLERRKAQLHQVANIL 60

Query: 69  RKDRDKYAEVMTKDMGKLFTEAQGRVDLCADIADYYADNQQKFLKPVPLESPNGEAYYLK 128
           R+DRDKYAR+MTKDMGKLFTEAQGRV+LCADIADYYAD   +FL   PLE+ +G+AYYLK
Sbjct: 61  RRDRDKYARIMTKDMGKLFTEAQGEVNLCADIADYYADKADEFLMSTPLETDSGQAYYLK 120

Query: 129 QAVGVLLAVEPWNFPFYQIMRVFAPNFIVGNTMLLKHASICPASAQAFEDLVREAGAPEG 188
           Q+ GV+LAVEPWNFP+YQIMRVFAPNFIVGN M+LKHASICP SAQ+FE+LV EAGA  G
Sbjct: 121 QSTGVILAVEPWNFPYYQIMRVFAPNFIVGNPMVLKHASICPRSAQSFEELVLEAGAPAG 180

Query: 189 APKNIFASYDQVSNLISDPRVAGVCLTGSERGGASIAAEAGKNLKKSSMELGGNDAFLIL 248
           +  N+F SYDQVS +I+D RV GVCLTGSERGGASIA EAGKNLKK++ELGG+DAF+IL
Sbjct: 181 SITNLFISYDQVSQVIADKRVVGVCLTGSERGGASIAEEAGKNLKKITLELGGDDAFIIL 240

Query: 249 DDADPDLLSKTIFFARLYNAGQVCTSSKRPIVMADKYDEFVNMVVETFKSAKWGDPMDSE 308
           DDAD+D L K ++F+RLYNAGQVCTSSKRPIV+   YD F  ++ + FK+AKWGDPMD E
Sbjct: 241 DDADWDQLEKVLYFSRLYNAGQVCTSSKRPIVLDKDYDRFKELITKVFKTAKWGDPMDPE 300

Query: 309 TTLAPLSSAGAKDDVLKQIKLAVDHGAEVVFGNDTIDHPGNFVMPTVLTNITKANPIYNQ 368
           TTLAPLSSA AK DVL QIKLA+DHGAE+V+G + IDHPG+FVMPT++  +TK NPIY Q
Sbjct: 301 TTLAPLSSAQAKADVLDQIKLALDHGAELVYGGEAIDHPGHFVMPTIIAGLTKDNPIYYQ 360

Query: 369 EIFGPVASIYKVDTREEAIALANDSSYGLGSTVFSSDPEHAKKVAAQIETGMTFINSGWT 428
           EIFGPV  IYKV +EEEAI +ANDS+YGLG T+FSS+ EHAK VAA+IETGM+FINSGWT
Sbjct: 361 EIFGPVGEIYKVSSEEEAIEVANDSNYGLGGTIFSSNQEHAKAVAAKIETGMSFINSGWT 420

Query: 429 SLPELPPGGIKNSGYGRELSQLGFDAFVNEHLVFTFN 465
           SLPELPPGGIK+SGYGRELS+LGF +FVNEHL++ FN
Sbjct: 421 SLPELPPGGIKHSGYGRELSELGFTSFVNEHLIYIFN 457
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1727

A DNA sequence (GBSx1832) was identified in *S.agalactiae* <SEQ ID 5367> which encodes the amino acid sequence <SEQ ID 5368>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1728

A DNA sequence (GBSx1833) was identified in *S.agalactiae* <SEQ ID 5369> which encodes the amino acid sequence <SEQ ID 5370>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.91   Transmembrane   94 - 110 (  86 - 115)
   INTEGRAL    Likelihood = -7.75   Transmembrane  154 - 170 ( 150 - 176)
   INTEGRAL    Likelihood = -7.11   Transmembrane  316 - 332 ( 312 - 339)
   INTEGRAL    Likelihood = -6.16   Transmembrane  258 - 274 ( 253 - 278)
   INTEGRAL    Likelihood = -2.71   Transmembrane  218 - 234 ( 217 - 234)
   INTEGRAL    Likelihood = -1.49   Transmembrane  286 - 302 ( 283 - 302)
   INTEGRAL    Likelihood = -0.96   Transmembrane   73 -  89 (  73 -  89)
   INTEGRAL    Likelihood = -0.27   Transmembrane  121 - 137 ( 121 - 137)

----- Final Results -----
              bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9861> which encodes amino acid sequence <SEQ ID 9862> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>gP:AAC75219 GB:AE000305 orf, hypothetical protein [Escherichia coli K12]
 Identities = 102/331 (30%), Positives = 172/331 (51%), Gaps = 26/331 (7%)

Query: 12  IPGLILCFIIA-IPSWLLGLYLPLIGAPVF-----AILIGIIVGSFVQNR--QLPNKGIA 63
           IPGL L  +I  +  W  G  +P +    F     AIL+G+++G+          +  + G+
Sbjct: 17  IPGLALSAVITGVALW--GGSIPAVAGAGFSALTLAILLGMVLGNTIYPHIWKSCDGGVL 74

Query: 64  FTSKYILQTAVVLLGFGLNLMQVMKVGISSLPYIIMTISISLLIAYVL-QKLFKLDKTIA 122
           F  +Y+L+  +L GF L   Q+  VGIS + I ++T+S + ++A  L QK+F LDK  +
Sbjct: 75  FAKQYLLRLGIILYGFRLTFSQIADVGISGIIIDVVITLSSTFLLACFLSQKVFGLDKHTS 134

Query: 123 TLIGVGSSICGGSAIAATAPVINAKDDEVAQAISVIFLFNILAALIFPTLGNFIG--LSD 180
            LIG GSSICG +A+ AT FV+ A+  +V  A++ + +F  +A  ++P +   +    S
Sbjct: 135 WLIGAGSSICGAAAVLATEPVVKAEASKVIVAVATVVIFGTVAIPLYPAIYPLMSQNFPSP 194

Query: 181 HGFALFAGTAVNDTSSVIAT--ATANDAINHSNTLGGATIVKLTFTLAIIPITIVLSIYH 238
            F ++ G+ V++ + V A   A + DA N       A I K+ R + + P  I+L+
Sbjct: 195 ETFGIYIGSTVHEVAQVVAAGHAISPDAEN------AAVISKMLRVMMLAPFLILLAA-R 247

Query: 239 MKQTQKEQSVSVTKI-PPKFVLYFILASLLTTIVASLGPSLRIPEPLKVLSKFFIVMAMG 297
           +KQ     S   +KI  P F + FI+ ++  +           +   L   L   F + MAM
Sbjct: 248 VKQLSGANSGRKSKITIPWFAILFIVVAIFNSPHL---LPQSVVNMLVTLDTFLLAMAMA 304

Query: 298 AIGINTNVSKLIKTGSKSILLGAACWLGIII 328
           A+G+ T+VS L K G K +L+   +I+
Sbjct: 305 ALGLTTHVSALKKAGAKPLLMALVLFAWLIV 335
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5371> which encodes the amino acid sequence <SEQ ID 5372>. Analysis of this protein sequence reveals the following:

```
        Possible site: 37
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.29   Transmembrane   30 -  46 (  22 -  50)
   INTEGRAL    Likelihood = -8.12   Transmembrane  314 - 330 ( 311 - 338)
   INTEGRAL    Likelihood = -6.05   Transmembrane    8 -  24 (   7 -  29)
   INTEGRAL    Likelihood = -6.00   Transmembrane  150 - 166 ( 146 - 172)
   INTEGRAL    Likelihood = -5.57   Transmembrane  257 - 273 ( 252 - 277)
   INTEGRAL    Likelihood = -3.50   Transmembrane   91 - 107 (  87 - 108)
   INTEGRAL    Likelihood = -2.60   Transmembrane   69 -  85 (  68 -  87)
   INTEGRAL    Likelihood = -2.55   Transmembrane  289 - 305 ( 289 - 305)
```

```
----- Final Results -----
                bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC75219 GB:AE000305 orf, hypothetical protein [Escherichia
coli]
 Identities = 100/329 (30%), Positives = 173/329 (52%), Gaps = 21/329 (6%)

Query: 8   LPGLLLCLLLALPAWCLGRLFPIIGAP----VFALLLGMLLA-LFYEWRDKTKEG-ISFT 61
           +PGL L  ++   A   G +  + GA        AILLGM+L    Y H  K+ +G + F
Sbjct: 17  IPGLALSAVITGVALWGGSIPAVAGAGFSALTLALLLGMVLGNTIYPHIWKGCDGGVLFA 76

Query: 62  SKYILQTAVVLLGFGLNLTQVNAVGMQSLPIIISTIATALLVAYGL-QKWLRLDVNTATL 120
            +Y+L+  ++L GF L  +Q+  VG+  + I + T+++  L+A  L QK   LD +T+ L
Sbjct: 77  KQYLLRLGIILYGFRLTFSQIADVGISGIIIDVLTLSSTFLLACFLGQKVFGLDKHTSWL 136

Query: 121 VGVGSSICGGSAVAATAPVIKAKDDEVAKAISVIFLPNMLAALLFPSLGQLLG--LSNEG 178
            +G GSSICG +AV AT PV+KA+  +V  A++ + +F  +A  L++P++  L+    S E
Sbjct: 137 IGAGSSICGAAAVLATEPVVKAEASKVTVAVATVVIFGTVAIFLYPAIYPLMSQMFSPET 196

Query: 179 FAIFAGTAVNDTSSVTATATAWDALHHSNTLDGATIVKLTRTLAILPITTLGLSLYRAKEE 238
           F I+ G+ V++ + V A   A       +  + A I K+ R + + P  + L+  R K+
Sbjct: 197 FGIYIGSTVHEVAQVVAAGHAIS----PDAENAAVISKMLRVMMLAPFLILLAA-RVKQL 251

Query: 239 HDIVTEENFSLRKSFPRFILFFLASLITTLMTSLGVSADSFHYLKTLSKFFIVMAMAAI 298
               + E   +  + P F + F++ ++  +       +  + L TL  F + MAMAA+
Sbjct: 252 SGANSGEKSKI--TIFWPAILFIVVAIPNSFHL---LPQSVVNMLVTLDTFFLAMAMAAL 306

Query: 299 GLNTNLVKLIKTGGCAILLGAI--CWVAI 325
           GL T++  L K G + +L+  +  W+ +
Sbjct: 307 GLTTHVSALKKAGAKPLLMALVLFAWLIV 335
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 225/333 (67%), Positives = 277/333 (82%), Gaps = 3/333 (0%)

Query: 11  KIPGLLLCFIIAIPSWLLGLYLPLIGAPVFALLIGIIVGSFYQNRQLFNNGIAFTSKYIL 70
           K+PGL+LC ++A+P+W LG   P+IGAPVFALL+G+++   FY++R    +GI+FTSKYIL
Sbjct: 7   KLPGLLLCLLLALPAWCLGRLPPIIGAPVFALLLGMLLALFYRHRDKTKEGISFTSKYIL 66

Query: 71  QTAVVLLGFGLNLMQVMKVGISSLPIIIMTISILIIAYVLQKLFKLDKTIATLIGVGSS 130
           QTAVVLLGFGLNL QVM VG+ SLPIII  TI+ +L++AY LQK  +LD   ATL+GVGSS
Sbjct: 67  QTAVVLLGFGLNLTQVMAVGMQSLPIIISTIATALLVAYGLQKWLRLDVNTATLVGVGSS 126

Query: 131 ICGGSAIAATAPVINAKDDEVAQAISVIFLPNILAALIFPTLGNFIGLSDHGFALFAGTA 190
           ICGGSA+AATAPVI AKDDEVA+AISVIFLPN+LAALLFP+LGQLLGLSNEGFAIFAGTA
Sbjct: 127 ICGGSAVAATAPVIKAKDDEVAKAISVIFLPNMLAALLFPSLGQLLGLSNEGFAIFAGTA 186

Query: 191 VNDTSSVTATATAWDAINHSNTLGGATIVKLTRTLAIIPITIVLSIYHMKQTQ---KEQS 247
           VNDTSSVTATATAWDA++HSNTL GATIVKLTRTLAI+PIT+ LS+Y  K+      E++
Sbjct: 187 VNDTSSVTATATAWDALHHSNTLDGATIVKLTRTIAILPITLGLSLYRAKKEHDIVTEEN 246

Query: 248 VSVTKIFPKFVLYFILASLLTTIVASLGFSLRIFEPLKVLSKFFIVMAMGAIGINTNVSK 307
            S+ K FP+F+L+F+LASL+TT++  SLG S   F  LK LSKFFIVMAM AIG+NTN+ K
Sbjct: 247 FSLRKSFPRFILFFLASLITTLMTSLGVSADSFHYLKTLSKFFIVMAMAAIGLNTNLVK 306

Query: 308 LIKTGGKSILLGAACWLGIIIVSLIMQAILGTW 340
           LIKTGG++ILLGA CW+ I  +VSL MQ  LG W
Sbjct: 307 LIKTGGQAILLGAICWVAITLVSLAMQLSLGIW 339
```

A related GBS gene <SEQ ID 8889> and protein <SEQ ID 8890> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:     22.17
GvH: Signal Score (-7.5): -0.429999
     Possible site: 41
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 8 value:  -7.91 threshold:  0.0
   INTEGRAL   Likelihood = -7.91   Transmembrane   94 - 110 (  86 - 115)
   INTEGRAL   Likelihood = -7.75   Transmembrane  154 - 170 ( 150 - 176)
   INTEGRAL   Likelihood = -7.11   Transmembrane  316 - 332 ( 312 - 339)
   INTEGRAL   Likelihood = -6.16   Transmembrane  258 - 274 ( 253 - 278)
   INTEGRAL   Likelihood = -2.71   Transmembrane  218 - 234 ( 217 - 234)
   INTEGRAL   Likelihood = -1.69   Transmembrane  286 - 302 ( 283 - 302)
   INTEGRAL   Likelihood = -0.96   Transmembrane   73 -  89 (  73 -  89)
   INTEGRAL   Likelihood = -0.27   Transmembrane  121 - 137 ( 121 - 137)
   PERIPHERAL Likelihood =  3.29     175
 modified ALOM score:   2.08

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02059(334 - 1284 of 1620)
RGAD|10465|EC2158(17 - 335 of 349) hypothetical 36.9 kd protein in lysp-nfo intergenic
region {Escherichia coli} OMNI|NT01EC2574 conserved hypothetical protein
SP|P33019|YSIH_ECOLI HYPOTHETICAL 36.9 KDA PROTEIN IN LYSP-NFO INTERGENIC REGION.
GP|405879|gb|AAA60511.1||U00007 yeiH {Escherichia coli} GP|1788482|gb|AAC75219.1||AE000305
orf, hypothetical protein {Escherichia coli} PIR|E64984|E64984 hypothetical 36.9 kD protein
in lysP-nfo intergenic region - Escherichia coli (strain K-12)
%Match = 12.7
%Identity = 32.3  %Similarity = 57.1
Matches = 103  Mismatches = 125  Conservative Sub.s = 79

270       300       330       360       390                   435       462
YSGPLSVFLSRFKACDIIVMVRRTIMLFKEKIPGLILCFIIAIPSWLLGLYLPLI-----GAPVFAILIGIIVG-SFYQN
                                  |||| |  :|     | |  :| :       |  :|||:|:::| : | :
MTNITLQKQHRTLWHFIPGLALSAVIT-GVALWGGSIPAVAGAGFSALTLAILLGMVLGNTIYPH
          10        20        30        40        50        60

489       519       549       579       609       636       666       696
R-QLFNKGIAFTSKYILQTAVVLLGFGLNLMQVMKVGISSLPIIIMTISISLIIAYVL-QKLFKLDKTIATLIGVGSSIC
  :  : |: |  :|:|:  ::| || | : |:  |||| : |::|:| ::::|  | |||: |||  : ||| |||||
IWKSCDGGVLFAKQYLLRLGIILYGFRLTFSQIADVGISGIIIDVLTLSGTFLLACFLGQKVFGLDKHTSWLIGAGSSIC
          80        90       100       110       120       130       140

726       756       786       816       840       870       900       930
GGSAIAATAPVINAKDDEVAQAISV-FLFNILAALIFPTLGNFIG--LSDHGFALPAGTAVNDTSSVTATATAWDAINHS
| :|: || ||: |:  :|  |:: : :|  :| :::| :  ::   :|   | :: |: |:: : |  |  ||
GAAAVLATEPVVKAEASKVTVAVATVVIFGTVAIFLYPAIYPLMSQWFSPETPGIYIGSTVHSVAQVVA---AGHAI-SP
          160       170       180       190       200       210       220

960       990      1020      1050      1077      1107      1134      1164
NTLGGATTVKLTTLAIIPITTVLSIYHMKQTQKEQSVSVTKI-FPKFVLYFILASLLTTIVASLGF-SLRIFEPLKVLS
:    | | |: | : : |  |:|:  :||     |  :||  | | : |:: :::     |: :   :   | |
DAENAAVISKMLRVMMLAPFLILLAA-RVKQLSGANSGKKGKITIPWFALILPIVVAIF----NSFHLLPQSVVNMLVTLD
        230       240       250       260       270       280       290

1194      1224      1254      1284      1314      1344      1374      1404
KFFIVMAMGAIGINTVSKLIKTGGKSILLGAACWLGIIIVSLTMQAILGTW*SCLKLNICNRFHKCYNEDIKRREHYGI
 |:: ||| |:|: |:|| | | | :|:   :  :|:
TFLLAMAMAALGLTTHVSALKKAGAKPLLMALVLFAWLIVGGSAINYVIQSVIA
          310       320       330       340
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1729

A DNA sequence (GBSx1834) was identified in *S.agalactiae* <SEQ ID 5373> which encodes the amino acid sequence <SEQ ID 5374>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.93   Transmembrane    7 -  23 (   1 -  27)

----- Final Results -----
                bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5375> which encodes the amino acid sequence <SEQ ID 5376>. Analysis of this protein sequence reveals the following:

```
        Possible site: 40

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-16.34   Transmembrane   22 -  38 (  13 -  42)

----- Final Results -----
                bacterial membrane --- Certainty=0.7538(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 56/215 (26%), Positives = 111/215 (51%), Gaps = 5/215 (2%)

Query: 7   VFLTVLVLILIVGAGGLYFWNNHQGLEGKWRTVGLEKQVEKEIEQQLGSQAADNGISAAD 66
           +F+ ++ LIL+    G+ + N+  S+EG WRT S+++++  +  ++L         I  +
Sbjct: 22  LFVFIIFLILLAVLFGVRYRNS--SIEGIWRTTSIDQKLGDDFAKRLTGLHQSPLIDDS- 78

Query: 67  LVKGANMHMNVKNDEAKITVTAQIDEVKFPHQAIKTFIDKALEKQLKDQGLTYNDLSEAGK 126
           L+  + M + VKN+   ++ + Q++   F + +  +   L K LK+  L   DLS     +
Sbjct: 79  LLTSSQMILTVKNNNVDLSFSVQVERDIFVKRLAAYHQNELLKTLKENNHLVVGDLSSKER 138

Query: 127 KIFDETKITDQQIDQQIDRSFQSAAQAAGGKYNTNTGEMTLPVMDGKVHRLTSVIKV-SH 185
           +I + +    +++  +D++F+  A   GGKYN  TG ++  V+ GKV+R+   I +
Sbjct: 139 QIIENSMPASHELEMILDQAFEKLASQIGGKYNQKTGHLSAVVLKGKVKRILHTIDIKEE 198

Query: 186 INKKANAFYGNIVKNGEKTAYKKEGSKL-ILGNKK 219
           +    +F  ++       Y + G KL +LG+EK
Sbjct: 199 VAAGHTSFSKGLLTPNGYFDYTRFGKKLELLGDEK 233
```

SEQ ID 5374 (GBS288) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 3; MW 53.7kDa).

GBS288d was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lane 8-10; MW 26kDa) and in Figure 183 (lane 3; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 187 (lane 11; MW 51kDa). Purified GBS288d-GST is shown in lane 8 of Figure 237.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1730

A DNA sequence (GBSx1835) was identified in *S.agalactiae* <SEQ ID 5377> which encodes the amino acid sequence <SEQ ID 5378>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3885(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1731

A DNA sequence (GBSx1836) was identified in *S.agalactiae* <SEQ ID 5379> which encodes the amino acid sequence <SEQ ID 5380>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.37   Transmembrane   67 -  83 (  63 -  89)
   INTEGRAL    Likelihood = -3.72   Transmembrane  139 - 155 ( 137 - 158)
   INTEGRAL    Likelihood = -1.54   Transmembrane  115 - 131 ( 114 - 131)

----- Final Results -----
               bacterial membrane --- Certainty=0.5946(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10905> which encodes amino acid sequence <SEQ ID 10906> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1732

A DNA sequence (GBSx1837) was identified in *S.agalactiae* <SEQ ID 5381> which encodes the amino acid sequence <SEQ ID 5382>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4709(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1733

A DNA sequence (GBSx1838) was identified in *S.agalactiae* <SEQ ID 5383> which encodes the amino acid sequence <SEQ ID 5384>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2191(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC98427 GB:M63481 20-kDa protein [Streptococcus sanguinis]
 Identities = 119/163 (73%), Positives = 146/163 (89%)

Query: 1   MTTFLGNPVTFTGKQLQVGDIAKDFLLIATDLSQKSLKDFEGKKKVISVVPSIDTGICSK 60
            MTTFLGNPVTFTGKQLQVGD A DF L ATDLS+K+L DF GKKKV+S++PSIDTG+CS
Sbjct: 1   MTTFLGNPVTFTGKQLQVGDTAHDFSLTATDLSKKTLADFAGKKKVLSIIPSIDTGVCST 60

Query: 61  QTRTFNEELSELDNTVVITVSMDLPFAQKRWCSAEGLDNVILLSDFYDHSFGQEYALLMN 120
            QTR FN+ELS+LDNTVVITVS+DLPFAQ +WC+AEG++N ++LSD++DHSFG++YA+L+N
Sbjct: 61  QTRRFNQELSDLDNTVVITVSVDLPFAQGKWCAAEGIENAVMLSDYFDHSFGRDYAVLIN 120

Query: 121 EWHLLTRAVLILDEHNKVTYTEYVDNVNSDVDYEAAINAAKIL 163
            EWHLL RAVL+LDE+N VTY EYVDN+N++ DY+AAI A K L
Sbjct: 121 EWHLLARAVLVLDENNTVTYAEYVDNINTEPDYDAAIAAVKSL 163
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1734

A DNA sequence (GBSx1839) was identified in *S.agalactiae* <SEQ ID 5385> which encodes the amino acid sequence <SEQ ID 5386>. This protein is predicted to be DNA alkylation repair enzyme. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4729(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB40581 GB:AJ010128 DNA alkylation repair enzyme [Bacillus
           cereus]
 Identities = 67/217 (30%), Positives = 119/217 (53%), Gaps = 5/217 (2%)
```

```
Query: 6    SLERKFKAASDKEVSKQQRAYLRHHFKCYGIKSPERRMLYKELIKAAKRQAKIDWQLLDK 65
             +L+  F A  + E ++     Y++HF   GI++PERR L K++I+       + D+Q++ +
Sbjct: 7    ALQEHFIANQNDEKAEPMARYMKNHFPFLGIQTPERRQLLKDVIQHTLPDQKDPQVIVR 66

Query: 66   -CWQSDYREYHHFVLDYLLAMSQFLTYNDCSRLEFYARHQQWWDSIDVLTKIF-GNLSLK 123
              W    RE+    LD +       +         LE     + WWD++D +    F GN+ L+
Sbjct: 67   ELWDLPERKFQAAALDMMQKYKMHINETHIPYLEELIVTKSWWDTVDSIVPTFLGNIFLQ 126

Query: 124  DDKVMNL-LSEWSLDQDFWMRRLAIEHQLGFKKETNTDILSLFILRNTCSQRFFINKAIG 182
             ++++   + +W     + W++R AI   QL  +K+K + ++L    I +   S+EFFI KAIG
Sbjct: 127  HPELISAYIPKWIASDNIWLQRAAILPQLKYKQKMDERLLPWVIGQLHSSKEFFIQKAIG 186

Query: 183  WALRDYSKYNKVWVKDFISNHCDRLSTLSIRRGSKYL 219
             W LR+Y+K     V +++ N  +EL+ LS RE  K++
Sbjct: 187  WVLREYAKTKSDVWKEYVQN--WELAPLSRRRAIKHI 221
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1735

A DNA sequence (GBSx1841) was identified in *S.agalactiae* <SEQ ID 5387> which encodes the amino acid sequence <SEQ ID 5388>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2117(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA81648 GB:Z27121 unknown [Mycoplasma hominis]
 Identities = 67/281 (23%), Positives = 113/281 (39%), Gaps = 52/281 (18%)

Query: 3    FVFDIDGTLCFDGMS--LSKEIQSILERAQIDYGHRVTFATARSYRDTIGILGDKLSLSK 60
             F  D+DGTL  D  +  +  + +  +++A +  GH V+  T R +R T+ +  +KL L+
Sbjct: 14   FAIDLDGTLLADSANGTVHPKTEEAIKKA-VAQGHIVSIITGRPWRSTLPVY-EKLGLNA 71

Query: 61   IIG-LNGATLHENGHLVDSYYLQSDFFSTIISYCHRHQIPYFVD------EVFNYATYQA 113
             I+G  NGA +H          FF   I+Y   +++ Y +          E+ NYA
Sbjct: 72   IVGNYNGAHIHNPA---------DPFFIPAITYLDLNEVLYILGDEKVKKEITNYAIEGP 122

Query: 114  SKIPFIAYVDPQ----------KRGELLEVSKIE----------KPIKMVLYFGDQLGR 152
              +  + + DP            K  E + + KI           KP   VL     L R
Sbjct: 123  DWYQLM-HRDPNLERVFGPNQATKFRECINLEKIPLKPTGIVFDVKPDTDVLELLTYLKR 181

Query: 153  ADQMLAELNRFGLSSHFPHREKCLYINPIAVDKGKATKKLFG------NRFIAPGNDKN 206
               L E + +          F+    I  I +DKGK     +      + +A G+  N
Sbjct: 182  RYGDLGEFSSWSKGKGLSPVFD----ITSIGIDKGKVISLIMRYYNIDIDDTVAMGDSYN 237

Query: 207  DISMFDAAHYSVQVGDPDELTPYANLRVSRESVHEGITTLF 247
             D+SM++ A+  V    + L    V +++   EG     F
Sbjct: 238  DLSMYNVANVCVSPANASPLIKKMSTVVMKQTNKEGAVGYF 278
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1736

A DNA sequence (GBSx1842) was identified in *S.agalactiae* <SEQ ID 5389> which encodes the amino acid sequence <SEQ ID 5390>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2383(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB90005 GB:AE001018 A. fulgidus predicted coding region AF1244
           [Archaeoglobus fulgidus]
 Identities = 22/48 (45%), Positives = 35/48 (72%)

Query: 150 GKSIGELNVWHQTGATIVAIEHEGKFIVSPGPFSVIEQGDHIFFVGDE 197
           GKSIGEL +  +TGAT++A+  + K I+SP P +V+E GD +  +G++
Sbjct: 102 GKSIGELGIRSKTGATVIAVLKKEKTIISPSPETVLSPGDKVVVIGEK 149
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5391> which encodes the amino acid sequence <SEQ ID 5392>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2446(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 163/213 (76%), Positives = 196/213 (91%)

Query: 1   MVSEQSEIVTSKYQKIAVAVAQRIANGDYEVGEKLKSRTTIASTFNVSPETARKGLNILA 60
           ++S + EI +SKYQKIA++VAQRIANG+YEVGEKLKSRTTIASTFNVSPETARKGLNILA
Sbjct: 1   VISPKKEITSSKYQKIAISVAQRIANGEYEVGEKLKSRTTIASTFNVSPETARKGLNILA 60

Query: 61  DLQILTLKHGSGAIILSKEKAIEPLNQYRTSHSVALLKGKIRDNIKAQQQEMEELATLVD 120
           DL+ILTLKHGSGAI+LSKE+AIEP+NQYR++HS+A+LK KIR+ I  Q + ME++A LV+
Sbjct: 61  DLKILTLKHGSGAIVLSKERAIRPINQYRSTHSIAVLKEKIRETINDQGKAMEKMAVLVN 120

Query: 121 DPLLQTRAVSKQYPLAPYEIIVSKDSEHLGKSIGELNVWHQTGATIVAIEHEGKFIVSPG 180
           DPL+Q++VSKQYPLAPYEII ++DSEH GKSIG LN+WHQTGATIVAIEH G+FIVSPG
Sbjct: 121 DPLMQSQSVSKQYPLAPYEIICNQDSEHPGKSIGVLNIWHQTGATIVAIEHAGQFIVSPG 180

Query: 181 PFSVIEQGDHIFFVGDEDVYARMKTYPNLRMGL 213
           P+SVIE+GDHI+FVGDE V +RMKT+PNLR GL
Sbjct: 181 PYSVIEKGDHIYFVGDESVISRMKTFPNLRKGL 213
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1737

A DNA sequence (GBSx1844) was identified in *S.agalactiae* <SEQ ID 5393> which encodes the amino acid sequence <SEQ ID 5394>. This protein is predicted to be gls24. Analysis of this protein sequence reveals the following:

```
Possible site: 16
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2855(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9361> which encodes amino acid sequence <SEQ ID 9362> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA86383 GB:U23376 putative 20-kDa protein [Lactococcus lactis]
 Identities = 63/124 (50%), Positives = 84/124 (66%)

Query: 1   MSGGFFSNLKNSVVNSDSVTDGVNVEVGTKEVAVDLDIVVEYGKDIPAIVSSIKAIVSQN 60
           + GGFFSNL   ++N+D VT GV+VEVG  +VAVDL +V EY K++P I E IK ++ +
Sbjct: 55  VEGGFFSNLTGKLINTDDVTTGVDVEVGKTQVAVDLKVVIEYRKNVPDIYEKIKEVIRKE 114

Query: 61  VEVMTHLKVVELNANVVDIKTKAEHEADSVTVQDRVSDAAQATGNFASEQAGKAKAAISS 120
           V  MT L+VVE+N  V DIKTK + + D V++QDRV+ AAQ TG F SEQ  K K  +
Sbjct: 115 VAAMTELEVVEVNVIVTDIKTKEQQKEDDVSIQDRVTSAAQTTGKFTSEQVDEVKDKVED 174

Query: 121 GAEK 124
            +K
Sbjct: 175 NTDK 178
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5395> which encodes the amino acid sequence <SEQ ID 5396>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2534(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 94/137 (68%), Positives = 108/137 (78%), Gaps = 8/137 (5%)

Query: 1   MSGGFFSNLKNSVVNSDSVTDGVNVEVGTKEVAVDLDIVVEYGKDIPAIVSSIKAIVSQN 60
           ++GGFFSN+KN++VNS+SVTDGV+VEVG+KEVAVDL I+VEYGKDIPAI ESIKAIVSQN
Sbjct: 35  VTGGFFSNIKNNLVNSESVTDGVSVEVGSKEVAVDLAIIVEYGKDIPAIAESIKAIVSQN 94

Query: 61  VEVMTHLKVVELNANVVDIKTKAEHEADSVTVQDRVSDAAQATGNFASEQAGKAKAAISS 120
           V+ MTHLKVVE+N NVVDI+TK EHEA SVTVQDRV+ AA +T  F SEQ  K K  IS
Sbjct: 95  VDSMTHLKVVEVNVVDIRTKEEHEHAASVTVQDRVTSAASSTSQFVSEQTEKLKDTISD 154

Query: 121 GAEKTKEAVSNGTEAAK 137
                    N  EAAK
Sbjct: 155 --------TVNSDEAAK 163
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1738

A DNA sequence (GBSx1845) was identified in *S.agalactiae* <SEQ ID 5397> which encodes the amino acid sequence <SEQ ID 5398>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3393(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1739

A DNA sequence (GBSx1846) was identified in *S.agalactiae* <SEQ ID 5399> which encodes the amino acid sequence <SEQ ID 5400>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3168(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1740

A DNA sequence (GBSx1847) was identified in *S.agalactiae* <SEQ ID 5401> which encodes the amino acid sequence <SEQ ID 5402>. This protein is predicted to be gls24. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2716(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA86383 GB:U23376 putative 20-kDa protein [Lactococcus lactis]
 Identities = 95/157 (60%), Positives = 120/157 (75%)

Query: 18  VRGELTFEDKVIRKIVGIAIEHVDGLLAVNGGFFSNLKNSVVNSDSVTDGVNVEVGKKQV 77
           ++G LT+EDKV++KIVG+A+E VDGLL+V GGFFSNL    ++N+D VT GV+VEVGK QV
Sbjct: 27  IKGALTYEDKVVQKIVGLALESVDGLLSVSGGFFSNLTGKLINTDDVTTGVDVEVGKTQV 86

Query: 78  AVDLDIVASYQKIVPTLFADIKGVVEAEVKRMTDLEVVSVNVNVVDIKTRAQHEEDSVTL 137
           AVDL +V EY+K+VP I+  IK+V+  EV  MT+LEVVSVNV V DIKT+ Q +ED V++
Sbjct: 87  AVDLKVVTSYRKNVPDIYEKIKSVIRKEVAAMTELEVVEVNVTVTDIKTKEQQKEDDVSI 146

Query: 138 QCRVTSAAQATGEFASNQVSNVKSAVGSGVDEVEDMK 174
           QCRVTSAAQ TG+F S QV  VK  V     DK   +K
```

```
Sbjct: 147 QDRVTSAAQTTGKFTSEQVDKVKDKVSDNFDKEARVK 183
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5403> which encodes the amino acid sequence <SEQ ID 5404>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3896(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 123/180 (68%), Positives = 158/180 (87%), Gaps = 1/180 (0%)

Query: 1   MTETYIKNTTNNSGTTAVRGELTFEDKVIEKIVGIAIEHVDGLLAVNGGFPSNLKNSVVN 60
           MTETYIKNT+ +  T+A+RG+LTF+DKVIEKIVG+A+E+VDGLL VNGGFP+NLK+ +VN
Sbjct: 1   MTETYIKNTSKDL-TSAIRGQLTYDDKVIEKIVGLALENVDGLLGVNGGFPANLKDKLVN 59

Query: 61  SDSVTDGVNVEVGKKQVAVDLDIVAEYQKHVPTIPADIKKVVEAEVKRMTDLEVVEVNVN 120
           ++SV DGVNVEVGKKQVAVDLDIVAEYQKHVPTI+  IK +VE EVKRMTDL+V+EVNV
Sbjct: 60  TESVRDGVNVEVGKKQVAVDLDIVAEYQKHVPTIYDSIKSIVEEEVKRMTDLDVIEVNVK 119

Query: 121 VVDIKTRAQHEEDSVTLQDRVTSAAQATGEFASNQVSNVKSAVGSGVDKVEDMKSEPRVQ 180
           VVDIKT+ Q E + V+LQD+V+  A++T EF S+QV NVK++V +GV+K++D K+EPRV+
Sbjct: 120 VVDIKTKEQFEAEKVSLQDKVSDMARSTSEFTSHQVENVKASVDNGVEKLQDQKAEPRVK 179
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1741

A DNA sequence (GBSx1848) was identified in *S.agalactiae* <SEQ ID 5405> which encodes the amino acid sequence <SEQ ID 5406>. This protein is predicted to be a 6-kDa protein. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL     Likelihood = -9.29   Transmembrane   25 -  41 (  23 -  52)

----- Final Results -----
               bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA86382 GB:U23376 putative 6-kDa protein [Lactococcus lactis]
 Identities = 27/61 (44%), Positives = 45/61 (73%)

Query: 3   EPVRKYRYPLGGAVIGLVLAAMLVTIGFFKTILALVIIVLGAYAGLYVQRTGMLDQFFNK 62
           ++  K RYP+ G ++G ++A  I TIGF+K IL L +I LG Y GL+++++G++DQF N+
Sbjct: 2   DYFEKNRYPIIGGIVGALIAVCIPTIGFWKMILVLFLIGLGIYIGLFLKKSGIIDQFINR 61

Query: 63 R 63
           +
Sbjct: 62 K 62
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5407> which encodes the amino acid sequence <SEQ ID 5408>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.73   Transmembrane   11 -  27 (   6 -  50)
   INTEGRAL    Likelihood = -7.11   Transmembrane   33 -  49 (  27 -  50)

----- Final Results -----
                bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 28/61 (45%), Positives = 48/61 (77%)

Query: 3  EFVRKYRYPLGGAVIGLVLAAMIVTIGPFKTILALVIIVLGAYAGLYVQRTGMLDQFFNKR 63
          EF  K++YP+ G ++GL++A +++  G FKT+LA++ I+LG Y GLY ++TG++DQF N++
Sbjct: 2  EFYEKFKYPIIGGLVGLIIIAILLMAFGLFKTLLAIIFIILGIYGGLYAKKTGIIDQFLNRK 62
```

A related GBS gene <SEQ ID 8891> and protein <SEQ ID 8892> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
MoG: Discrim Score:    12.56
GvH: Signal Score (-7.5): -1.11
     Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value:  -9.29 threshold:  0.0
   INTEGRAL    Likelihood = -9.29   Transmembrane   25 -  41 (  23 -  52)
   PERIPHERAL  Likelihood = 12.25       44
 modified ALOM score:   2.36

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
44.3/73.8% over 60aa
  Lactococcus lactis
  EGAD|42618| putative 6-kDa protein Insert characterized
  GP|727435|gb|AAA86382.1||U23376 putative 6-kDa protein Insert characterized

ORF01006(307 - 489 of 792)
EGAD|42618|45008(2   -    62    of    62)    putative    6-kDa    protein    (Lactococcus
lactis)GP|727435|gb|AAA86382.1||U23376 putative 6-kDa protein (La
ctococcus lactis)
%Match = 11.6
%Identity = 44.3  %Similarity = 73.8
Matches = 27  Mismatches = 16  Conservative Sub.s = 18

159       189       219       249       279       309       339       369
TNVPEQLEHIQSDVELGLKEFPGLSKKMNTRVFVKQVEENVGNAKTNKSRVE*ESNMSEFVRKYRYPLGGAVIGLVLAA
                                                      ::  | |||: | ::| ::|
                                                    MDYFKKNRYPIIGGIVGALIAV
                                                             10        20

399       429       459       489       519       549       579       609
MIVTIGFFKTILALVIIVLGAYAGLYVQRTGMLDQFFNKRK*NFSFIFILHIYLNKRRRNYYD*NLHQKHN*QFWHDSCSW
 | ||||:| || | :| || | ||:::::|::||| |::
CIFTIGFWKMILVLFLIGLGIYIGLWLKKSGIIDQFINRK
                  40        50        60
```

SEQ ID 5406 (GBS14) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 9 (lane 4; MW 33.3kDa). The GBS14-GST fusion product was purified (Figure 190, lane 8) and used to immunise mice. The resulting antiscrum was used for FACS (Figure 263), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1742

A DNA sequence (GBSx1849) was identified in *S.agalactiae* <SEQ ID 5409> which encodes the amino acid sequence <SEQ ID 5410>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-18.63   Transmembrane   61 -  77 (  51 -  83)
   INTEGRAL    Likelihood = -7.91   Transmembrane   10 -  26 (   7 -  28)

----- Final Results -----
                bacterial membrane --- Certainty=0.8451(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5411> which encodes the amino acid sequence <SEQ ID 5412>. Analysis of this protein sequence reveals the following:

```
      Possible site: 29

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-16.19   Transmembrane   71 -  87 (  63 -  93)

----- Final Results -----
                bacterial membrane --- Certainty=0.7474(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 87/193 (45%), Positives = 127/193 (65%), Gaps = 4/193 (2%)

Query: 1   MSKGLKSLYTLLGLISLTLLGFVAVISKQHIVLP-SFNWLEWDFN-LPSPIDVGMYHYFF 58
           MSK LK  Y L+GL+ L++ G+V  I+  +IYLP S+ WL W  +  P+ +D  + +Y+F
Sbjct: 9   MSKLLKISYCLVGLVLLSVFGWVVGITGGYIYLPYSYRWLSWGWDSFPNLLDSALSYYYF 68

Query: 59  WCALVLFVIVLLAILVVLFYPRRYTEYKLA--DKTGKLMLKKSAIEGFVKTEVLKTGLMK 116
           W ALVLFVI  LA+LVV++ YPR YTE +L    +K G L+LKKSAIE +V T +   GLM
Sbjct: 69  WTALVLFVITPLALLVILYPRIYTKVQLRHKNKKGTLILLKKSAIESYVATAIQTAGLMP 128

Query: 117 SPSVIAHLYKKKVKVDVKGLLTSRTNVPEQLEHIQSDVELGLKEFPGLSKKMETRVFVKQ 176
           +P+VTA LYK+K  + VKG L SR  V +Q+  ++  +E GL EFPG+    +N +V+VK
Sbjct: 129 NPTVTAKLYKRKFNIIVKGRLASRVAVADQISGVKEGIEKGLTEFPGINYPVNFKVYVKD 188

Query: 177 VEKENVGNAKTNK 189
           + + +  +   N+
Sbjct: 189 IADSDRKHITTNR 201
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1743

A DNA sequence (GBSx1850) was identified in *S.agalactiae* <SEQ ID 5413> which encodes the amino acid sequence <SEQ ID 5414>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.82   Transmembrane   56 -  72 (  52 -  81)
   INTEGRAL    Likelihood = -6.42   Transmembrane    4 -  20 (   1 -  23)

----- Final Results -----
                bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12244  GB:Z99106  similar  to  hypothetical  proteins  from  B.  subtilis  [Bacillus
subtilis]
 Identities = 31/76 (40%), Positives = 48/76 (62%)

Query: 1  MSLIWSLIVGAIIGAIAGAVTNKGGSMGWIANILAGLVGSFVGQSLLGTWGPKLAGMALI 60
          +S + SL+V  +IG I  A+       G   +++AGL+G+++G  LLGTWGP LAG A+
Sbjct: 2  LSFLVSLVVAIVIGLIGSAIVGNRLPGGIFGSMIAGLIGAWIGHGLLGTWGPSLAGFAIF 61

Query: 61 PSIVGAIIVVIVTSFV 76
           P+I+GA I V +   +
Sbjct: 62 PAIIGAAIFVFLLSLI 77
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5415> which encodes the amino acid sequence <SEQ ID 5416>. Analysis of this protein sequence reveals the following:

```
      Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.59   Transmembrane   60 -  76 (  56 -  80)

----- Final Results -----
                bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12244  GB:Z99106  similar  to  hypothetical  proteins  from  B.  subtilis  [Bacillus
subtilis]
 Identities = 28/76 (36%), Positives = 47/76 (61%)

Query: 1  MGLIWTLIVGALIGVIAGALVTKKGGSMGWIANIAAGLVGSSVGQALLGSWGPSLAGMSLI 60
          +  + +L+V  +IG+I  A+       G   ++ AGL+G+ +G  LLG+WGPSLAG  ++
Sbjct: 2  LSFLVSLVVAIVIGLIGSAIVGNRLPGGIFGSMIAGLIGAWIGHGLLGTWGPSLAGFAIF 61

Query: 61 PSVIGAVIVVMITSFV 76
           P++IGA I V +   +
Sbjct: 62 PAIIGAAIFVFLLGLI 77
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 63/82 (76%), Positives = 74/82 (89%)

Query: 1  MSLIWSLIVGAIIGAIAGAVTNKGGSMGWIANILAGLVGSFVGQSLLGTWGPKLAGMALI 60
          M LIW+LIVGA+IG IAGA+T KGGSMGWIANI AGLVGS VGQ+LLG+WGP LAGM+LI
Sbjct: 1  MGLIWTLIVGALIGVIAGALVTKKGGSMGWIANIAAGLVGSSVGQALLGSWGPSLAGMSLI 60

Query: 61 PSIVGAIIVVIVTSFVLGKMNN 82
           PS++GA+IVV++TSFVL K NN
Sbjct: 61 PSVIGAVIVVMITSFVLNKTNN 82
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1744

A DNA sequence (GBSx1851) was identified in *S.agalactiae* <SEQ ID 5417> which encodes the amino acid sequence <SEQ ID 5418>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.82   Transmembrane   88 - 104 (  84 - 111)
   INTEGRAL    Likelihood = -8.07   Transmembrane   29 -  45 (  27 -  54)

----- Final Results -----
               bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12244 GB:Z99106 similar to hypothetical proteins from B. subtilis [Bacillus
subtilis]
 Identities = 29/77 (37%), Positives = 47/77 (60%)

Query: 31  IMGLIWSLIVGAIIGAIAGAITNKGGSMGWIANILAGLVGSFVGQSLLGTWGPKLADMAL 90
           ++  + SL+V  +IG I  AI       G   +++AGL+G+++G  LLGTWGP LA  A+
Sbjct: 1   MLSFLVSLVVAIVIGLIGSAIVGNRLPGGIFGSMIAGLIGAWIGHGLLGTWGPSLAGFAI 60

Query: 91  IPSIVGAIIVIIVTSFV 107
            P+I+GA I + +   +
Sbjct: 61  FPAIIGAAIFVFLLGLI 77
```

There is also homology to SEQ ID 5416:

```
Identities = 60/79 (75%), Positives = 72/79 (90%)

Query: 32  MGLIWSLIVGAIIGAIAGAITNKGGSMGWIANILAGLVGSFVGQSLLGTWGPKLADMALI 91
           MGLIW+LIVGA+IG IAGA+T KGGSMGWIANI AGLVGS VGQ+LLG+WGP LA M+LI
Sbjct: 1   MGLIWTLIVGALIGVIAGALTKKGGSMGWIANIAAGLVGSSVGQALLGSWGPSLAGMSLI 60

Query: 92  PSIVGAIIVIIVTSFVLGK 110
           PS++GA+IV+++TSFVL K
Sbjct: 61  PSVIGAVIVVMITSFVLNK 79
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1745

A DNA sequence (GBSx1852) was identified in *S.agalactiae* <SEQ ID 5419> which encodes the amino acid sequence <SEQ ID 5420>. This protein is predicted to be ATP-dependent DNA helicase Rep (uvrD). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1364(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9863> which encodes amino acid sequence <SEQ ID 9864> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD51119 GB:AF176554 DNA helicase PcrA [Leuconostoc citreum]
 Identities = 414/764 (54%), Positives = 537/764 (70%), Gaps = 23/764 (3%)

Query: 6   VEMNPLIIGMNDKQARAVQTTDGPLLIMAGAGSGKTRVLTHRIAYLIDEKYVNPWNILAI 65
           + +  L  GMN+KQAEAVQTT+GPLLIMAGAGSGKTRVLTHRIA+L+ +  V PW ILAI
Sbjct: 1   MSVETLTNGMNNKQAEAVQTTEGPLLIMAGAGSGKTRVLTHRIAHLVQDLNVFPWRILAI 60

Query: 66  TFTNKAAREMRERAIAL--NPATQDTLIATFHSMCVRILRREADYIGYNRNFTIVDPGEQ 123
           TFTNKAAREMRER AL      +D  ++TFH++ VRILRR+ + IG  +NFTI+D   Q
Sbjct: 61  TFTNKAAREMRERIAALLSEDVARDIWVSTFHALAVRILRRDGHAIGLAKNFTIIDTSAQ 120

Query: 124 RTLMKRIIKQLNLDTKKWNERSILGTISNAKNDLLDEIAYEKQAGDMYTQVIAKCYKAYQ 183
           RTLMKR+I  LNLDT +++ R+ILG ISNAKND+L    Y K A + + + +A+ Y AYQ
Sbjct: 121 RTLMKRVINDLNLDINQYDPRTILGMISNAKNDMLQPRDYAKAADNAFQETVAEVYTAYQ 180

Query: 184 EELRRSEAMDFDDLIMWTLRLFDQNKDVLAYYQQRYQYIEVDEYQDTNHAQYQLVKLLAS 243
            EL+RS++DFDDLIM+T+ LF    DVLA YQQ+++Y+EVDEYQDTN AQY +V LLA
Sbjct: 181 AELKRSQSVDFDDLIMLTIDLFQSADPVLARYQQQFEYLHVDEYQDTNDAQYTIVNLLAQ 240

Query: 244 RFKNICVVGDADQSIYGWRGADMQNILDFEKDYPQAKVVLLEENYRSTKKILQAANNVIN 303
           R KN+ VVGDADQSIYGWRGA+M NIL+FEKDYP A  V+LE+NYRST+ IL AAN VIN
Sbjct: 241 RSKNLAVVGDADQSIYGWRGANMNNILNFEKDYPNAHTVMLEQNYRSTQNILDAANAVIN 300

Query: 304 HNKNRRPKKLWTQNDEGBQIVYHRANNEQEEAVFVASTIDNIVREQGKNFKDFAVLYRTN 363
           HN  R PKKLWT+N +G+QI Y+RA  E +EA F+ S I  +   +  DFAVLYRTN
Sbjct: 301 HNNERVPKKLWTENGKGDQITYYRAQTEHDEANFILSNIQQLRSTKHMAYSDFAVLYRTN 360

Query: 364 AQSRTIEEALLKSNIPYTMVGGTKFYSRKEIRDVIAYLNILANTSDNISFERIVNEPKRG 423
           AQSR IEE+LL+K+N+PY+MVGG KPY RKEI D++AY+++++ N  DN +FER+VNEPKRG
Sbjct: 361 AQSRNIEESLVKANMPYSMVGGHKPYERKEILDIMAYMSLITNPDDNAAPERVVNEPKRG 420

Query: 424 VGPGTLEKIRSPAYEQSMSLLDASSNVMMSP-LKGKAAQAVWDLANLILTLRSNLDSLTV 482
           +G  +L ++R  A  ++S + A  ++ ++P +  KAA      A ++  LR   + L V
Sbjct: 421 LGATSLTRLRELANRLNVSYMKAIGSIELAPSITTKAASKFLTFAEMMHNLRQQSEFLNV 480

Query: 483 TEITENLLDKTGYLEALQVQNTLESQARIENIEEPLSVTNFDDNPEITVEGETGLDRLS 542
           TE+TE ++ ++GY + L  +N  +SQAR+EN+EEPLSVTK FDD  +   E  +D ++
Sbjct: 481 TELTELVMTQSGYRQMLAEKNDFDSQARLENLSEFLSVTKEFDD--KYQPEDPESIDPVT 538

Query: 543 RFLNDLAIDATDDSATETAEVTLMTLHAAKGLEPPVVFLIGMEEGVFPLSRAIEDADEL 602
            PL  AL++D DD        VTLMTLHAAKGLEPPVVFLIG++EGVFPLSRAI+ D D L
Sbjct: 539 DPLGTTALMSDLDDFEEGDGAVTLMTLHAAKGLEPPVVFLIGLKEGIFPLSRAMDEDLL 598

Query: 603 EEERRLAYVGITRAKQILFLTNANIRTLFGKTSYNRPTRFIREIDDELIQ--YQGLARPV 660
           EEERRLAYVGITRA + LFLTNA +R L+G+T  N P+RFI EI   EL++  Y GL+R
Sbjct: 599 EEERRLAYVGITRAMKKLFLTNAFSRLLYGRTQANEPSRFIAEISPELLETAYSGLSRDK 658

Query: 661 NSSPGVKYSKRQPTQPGQGMSLQQALQARKSNSQSQVIAQLQALN-ANNSHETSWRIGDV 719
           + + ++          R + + Q T  +  N     +TSW  GD
Sbjct: 659 TQKTLPFDRK----------MQRATATTYQATFVTKITNGVTGDQTSWSTGDK 703

Query: 720 ATHKKWGDGTVLEVESGSGKTQRLKINFPGIGLKKLLASVAPISK 763
            +HKKWG GTV+ VSG     QRLK+ FP  G+K+LLA+ API K
Sbjct: 704 VSHKKWGVGTVISVSGRADDQELKVAFPSEGVKQLLAAFAPIQK 747
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5421> which encodes the amino acid sequence <SEQ ID 5422>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
                bacterial cytoplasm --- Certainty=0.0214(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 622/772 (80%), Positives = 699/772 (89%), Gaps = 15/772 (1%)

Query: 8    MNPLIIGMNDKQASAVQTTDGPLLIMAGAGSGKTRVLTHRIAYLIDEKYVNPWNILAITF 67
            MNPL+ GMND+QA+AVQTT+GPLLIMAGAGSGKTRVLTHRIAYLIDEK+VNPWNILAITF
Sbjct: 1    MNPLLNGMNDRQAQAVQTTEGPLLIMAGAGSGKTRVLTHRIAYLIDEKFVNPWNILAITF 60

Query: 68   TNKAAREMRERALALNPATQDTLIATPHSMCVRLLRREADYIGYNRNFTIVDPGEQRTLM 127
            TNKAAREM+ERA+ALNPAT+DTLIATPHSMCVRLLRREAD+IGYNRNFTIVDPGEQRTLM
Sbjct: 61   TNKAAREMKERALALNPATKDTLIATPHSMCVRLRREADHIGYNRNFTIVDPGEQRTLM 120

Query: 128  KRIIKQLNLDTKKWNERSILGTISNAKNDLLDEIAYEKQAGDMYTQVIAKCYKAYQEELR 187
            KRI+KQLN+D KKWNERSILGTISNAKNDLLDE  Y$ QA DMY+Q++A+CYKAYQEELR
Sbjct: 121  KRILKQLNIDPKKWNKRSILGTISNAKNDLLDEKGYEAQAADMYSQIVARCYKAYQEELR 180

Query: 188  RSEAMDFDDLIMMTLRLFDQNKDVLAYYQQRYQYIHVDEYQDTNHAQYQLVKLLASRFKN 247
            RSEA+DFDDLIMMTLRLFD N DVLAYYQQRYQYIHVDEYQDTNHAQYQL+KLLASRFKN
Sbjct: 181  RSEALDFDDLIMMTLRLFDANPDVLAYYQQRYQYIHVDEYQDTNHAQYQLIKLLASRFKN 240

Query: 248  ICVVGDADQSIYGWRGADMQNILDFEKDYPQAKVVLLEENYRSTKKILQAANNVINNKNN 307
            ICVVGDADQSIYGWRGADMQNILDFEKDYP AKVVLLEENYRSTKKILQAAN+VIN+N+N
Sbjct: 241  ICVVGDADQSIYGWRGADMQNILDFEKDYPDAKVVLLEENYRSTKKILQAANDVINNNRN 300

Query: 308  RRPKKLWTQNDEGEQIVYHRANNEQEEAVFVASTIDNIVREQGKNFKDFAVLYRTNAQSR 367
            RRPKKLWTQN +GEQ+VY+RAN+E++EAVFVASTI N+ +E GKNFKDFAVLYRTNAQSR
Sbjct: 301  RRPKKLWTQNADGEQLVYYRANDERDEAVFVASTISNMSQELGKNFKDFAVLYRTNAQSR 360

Query: 368  TIEEALLKSNIPYTMVGGTKFYSRKEIRDVIAYLNILANTSDNISFERIVNEPKRGVGPG 427
             TIEEALLKSNIPYTMVGGTKFYSRKEIRD+IAYL I+AN +DNISFERIVNEPKRGVGPG
Sbjct: 361  TIEEALLKSNIPYTMVGGTKFYSRKEIRDLIAYLITVANPADNISFERIVNEPKRGVGPG 420

Query: 428  TLEKIRSFAYSQGMSLLDASSNVMMSPLKGKAAQAVWDLANLILTLRSNLDSLTVTEITE 487
             TL+K+R FAYE   SLL+A+SN++MSPLKGKAAQA+ DLAN++  LR +LD +++T++ E
Sbjct: 421  TLDKLRQFAYESSDQSLLEAASNLLMSPLKGKAAQAIMDLANILGQLRQDLDQMSITDLAE 480

Query: 488  NLLDKTGYLEALQVQNTLESQARIENIEEFLSVTKNFDDNPEITVEGETGLDRLSRFLND 547
              LL+KTGYL++L++QNTLESQARIENIEEFLSVTKNFD++      E ETG+DRL RFLND
Sbjct: 481  ALLEKTGYLDSLRIQNTLESQARIENIEEFLSVTKNFDESSASQEEDETGVDRIGRFLND 540

Query: 548  LALIADTDDSATETAEVTLMTLHAAKGLEFPVVFLIGMEEGVFPLSRAIEDADELEEERR 607
             LALIADTDDS  E AEVTLMTLHAAKGLEFPVVFLIGMEEGVFPLSRA ED DELEEERR
Sbjct: 541  LALIADTDDSQAEAAEVTLMTLHAAKGLEFPVVFLIGMEEGVFPLSRASEDDPDELEEERR 600

Query: 608  LAYVGITRAEQILFLTNANTRTLFGKTSYNRPTRFIREIDDELIQYQGLARPVNESSPGVK 667
             LAYVGITRAE++LF+TNANTRTLFGK+SYNRPTRF++EI +EL+ Y+GLARP  SSPGV+
Sbjct: 601  LAYVGITRAEVLFMTNANTRTLFGKSSYNRPTRFLKEISEELLSYKGLARPAQSSPGVR 660

Query: 668  YSKEQPTQFGQGMSLQQALQARKSNSQSQVTAQ-LQA----------LNANNS-HET 712
             +S E  TQFGQGMSL +ALQARK+ +Q + +AQ +QA                +N+S  E
Sbjct: 661  FSTETHTQFGQGMSLSKALQARKAQAQVRQSAQPMQAHIIPSASTSSVLPFGSNSSVEEV 720

Query: 713  SWEIGDVATHKKWGDGTVLEVSGSGKTQELKINFPGIGLKKLLASVAPISKK 764
             +W+IGD+A HKKWGDGTVLEVSGSGKT ELKI FP +GLKKLLASVAPI KK
Sbjct: 721  TWQIGDIAHHKKWGDGTVLEVSGSGKTMELKIKFPEVGLKKLLASVAPIKKK 772
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1746

A DNA sequence (GBSx1853) was identified in *S.agalactiae* <SEQ ID 5423> which encodes the amino acid sequence <SEQ ID 5424>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4741(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA88579 GB:M14339 unknown [Streptococcus pneumoniae]
 Identities = 43/57 (75%), Positives = 50/57 (87%)

Query: 41 AHGGYLFTLCDQVSGLVAISTGYEAVTLQSNINYLRAGRLDDLLTVIGTCVHNGRTT 97
          AHGGYLFTLCDQ+SGLV IS G + VTLQS+INYL+AG+LDD+LT+ G CVH GRTT
Sbjct: 1  AHGGYLFTLCDQISGLVVISLGLDGVTLQSSINYLKAGKLDDVLTIKGECVHQGRTT 57
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5425> which encodes the amino acid sequence <SEQ ID 5426>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1210(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 57/97 (58%), Positives = 74/97 (75%)

Query: 2   KFNLEQVKVFENYEIENWEEGQVTLLTKVVDSSLNYYGNAHGGYLFTLCDQVSGLVAIST 61
           +  L   + +F+NY+IE  E+G + L+T+V +++LNYYGNAHGGYLFTLCDQV GLVA +T
Sbjct: 7   EMTLNVISIFDNYQIELAEKGHLILSTEVTETALNYYGNAHGGYLFTLCDQVGGLVARTT 66

Query: 62  GYEAVTLQSNINYLRAGRLDDLLTVIGTCVHNGRTTK 98
           G E+VTLQ+N NYL+AG   D L V G  VH GRTT+
Sbjct: 67  GVESVTLQANANYLKAGHKDKLMVEGRLVHGGRTTQ 103
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1747

A DNA sequence (GBSx1854) was identified in *S.agalactiae* <SEQ ID 5427> which encodes the amino acid sequence <SEQ ID 5428>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3187(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1748

A DNA sequence (GBSx1855) was identified in *S.agalactiae* <SEQ ID 5429> which encodes the amino acid sequence <SEQ ID 5430>. This protein is predicted to be uracil permease (uraA). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.65   Transmembrane  122 - 138 ( 117 - 146)
   INTEGRAL    Likelihood = -8.65   Transmembrane  212 - 228 ( 204 - 233)
   INTEGRAL    Likelihood = -7.32   Transmembrane   60 -  76 (  49 -  80)
   INTEGRAL    Likelihood = -6.53   Transmembrane  149 - 165 ( 145 - 172)
   INTEGRAL    Likelihood = -6.48   Transmembrane  402 - 418 ( 401 - 420)
   INTEGRAL    Likelihood = -4.04   Transmembrane  422 - 438 ( 420 - 445)
   INTEGRAL    Likelihood = -3.72   Transmembrane  365 - 381 ( 364 - 385)
   INTEGRAL    Likelihood = -3.40   Transmembrane  184 - 200 ( 182 - 202)
   INTEGRAL    Likelihood = -3.08   Transmembrane  346 - 362 ( 345 - 363)
   INTEGRAL    Likelihood = -1.38   Transmembrane  260 - 276 ( 260 - 276)

----- Final Results -----
               bacterial membrane --- Certainty=0.4661(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9865> which encodes amino acid sequence <SEQ ID 9866> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA53697 GB:X76083 uracil permease [Bacillus caldolyticus]
 Identities = 208/416 (50%), Positives = 291/416 (69%), Gaps = 11/416 (2%)

Query: 32  LLDIDEKPSLPQGLLLSPQHVFAMPGATILVPLILGMPVSVALFASGCGTLIYQVATKPK 91
           +LDI ++P + Q + LS QH+FAMPGATILVP ++G+  S+AL  SG GTL + + TK++
Sbjct: 5   VLDIQDRPTVGQWITLSLQHLFAMPGATILVPYLVGLDPSIALLTSGLGTLAFLLITKMQ 64

Query: 92  VPVYLGSSFAYITAMALAMKQMHGDISAAQTGILPVGLIYVVVATVIKPVGNSWVDKILP 151
           VP YLGSSFAYI  + A  +  G   AA  G    GL+Y VVA +IK  G  WV K+LP
Sbjct: 65  VPAYLGSSFAYIAPIIAA--KTAGGPGAAMIGSPLAGLVYGVVALIIKKAGYRWVMKLLP 122

Query: 152 PIIIGPMIIVIGLGLANSAVTNA--GPVAKGDWRKMLVAVVTPLIAAPINTKGKGFIKII 209
           P+++GP+IIVIGLGLA +AV  A  G   K        VA+VT       +  +G + +I
Sbjct: 123 PVVVGPVIIVIGLGLAGTAVGMAMNGPDGKYSLLHFSVALVTLAATIVCSVLARGMLSLI 182

Query: 210 PFLFAIIGGYILSIILGLVDLSPVEKAANFFLPKFYLPFKTGLFHSYKLYPGPEMLAIL- 268
           P L  I+ GY+ ++ +GLVDLS V  A WFE P F +PF       Y +     E++ ++
Sbjct: 183 PVLVGIVVGVLYALAVGLVDLSKVAAAKWFEWPDFLIPFA-----DYPVRVTWEVWLMV 237

Query: 269 PISIVTIAENIGDHTVLGQICGRNFLKKPGLNRLLIGDGLATAFSALIGGPAETTYGENT 328
           P++IVT++E+IG   VL ++ GR+ ++KPGL+R ++GDG AT  SAL+GGP +TTYGEN
Sbjct: 238 PVAIVTLSEHIGHQLVLSKVVGRDLIQKPGLHRSILGDGTATMISALLGGPPKTTYGENI 297

Query: 329 GVIGMTRIASVTVIRNAAFIAIAPSFPGKFTALISTIPSAVLGGMAILLYGVIASNGLKV 388
           GV+ +TR+ SV V+  AA IAIAP F GK TALIS+IP+ V+GG++ILL+G+IAS+GL++
Sbjct: 298 GVLAITRVYSVYVLAGAAVIAIAPGFVGKITALISSIPTPVMGGVSILLFGIIASSGLRM 357

Query: 389 LIENRVNPARVRNLIIASSMLVLGLGGAVLDLG-ALTLSGTALSAIVGIILNLLLP 443
           LI++RV+P + RNL+IAS +LV+G+GGAVL +  ++G ALGAIVG++LNLLLP
Sbjct: 358 LIDSRVDPGQTRNLVIASVILVIGIGGAVLKISDSFQITGMALSAIVGVLLNLLLP 413
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5431> which encodes the amino acid sequence <SEQ ID 5432>. Analysis of this protein sequence reveals the following:

```
        Possible site: 27
    >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood =-10.72   Transmembrane  177 - 193 ( 171 - 206)
    INTEGRAL    Likelihood = -8.55   Transmembrane  313 - 329 ( 304 - 339)
    INTEGRAL    Likelihood = -8.17   Transmembrane  154 - 170 ( 152 - 175)
    INTEGRAL    Likelihood = -7.91   Transmembrane  376 - 392 ( 374 - 395)
    INTEGRAL    Likelihood = -7.48   Transmembrane   25 -  41 (  22 -  43)
    INTEGRAL    Likelihood = -5.84   Transmembrane  120 - 136 ( 116 - 142)
    INTEGRAL    Likelihood = -4.99   Transmembrane   96 - 112 (  90 - 117)
    INTEGRAL    Likelihood = -3.29   Transmembrane  339 - 355 ( 338 - 360)
    INTEGRAL    Likelihood = -1.91   Transmembrane  396 - 412 ( 396 - 413)

    ----- Final Results -----
                bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
    >GP:CAB89870 GB:AJ132624 uracil transporter [Lactococcus lactis]
     Identities = 294/421 (69%), Positives = 359/421 (84%), Gaps = 5/421 (1%)

    Query: 3   DVIYDVEEVPKAGMLVGLSPQHLFAMFGATVLVPILVGIDPSVALLSSGLGTLAHLSVTK 62
              D+I  V+E P A    GLSPQHLFAMFG+TVLVPILVGI+P++ALLSSGLGTLAH+SVTK
    Sbjct: 5   DIILKVDEKPAASQWPGLSPQHLFAMFGSTVLVPILVGINPALALLSSGLGTLAHMSVTK 64

    Query: 63  FKIPAYMGSSFAYIAAMQLLMKTNGIGAVAQGAMTGGLVYLIVALIVKAIGNDWIDNILP 122
              FK+PAYMGSSFAYI AM LLMK  G+ A+AQGAMTGGLVYLIVALIVK  G  WID +LP
    Sbjct: 65  FKVPAYMGSSFAYIGAMTLLMKNGGMPAIAQGAMTGGLVYLIVALIVKFAGKGWIDKVLP 124

    Query: 123 PIVVGPIVMVIGLSLASTAVNDVMLKN----GNYNLTYLVIGLVTLLSVIFFNIYGKGIV 178
              PIVVGPIVMVIGLSLA TA+ND M  +      Y+L Y++I L+T+LS++ ++IYGKG +
    Sbjct: 125 PIVVGPIVMVIGLSLAPTAINDAMYTDVANLKGYSLAYIIIALITVLSIVVSIYGKGFL 184

    Query: 179 AIVPLLLGLLVGYVVALLVGVLTGQEIVDFTNVAQAKWPSIPSVRIPFLTYGVKFYPSAI 238
              ++VP+LLG++ GYV A++G +TG  IV FT ++QAKW ++P +EIPF +Y   FYPSAI
    Sbjct: 185 SVVPILLGIITGYVAAMIIGKITGMNIVSFTGISQAKWLTLPPMEIPFASYKWAPYPSAI 244

    Query: 239 LTMAPIAFVTMTEHPGHIMVLNSLTKRDYFKDPGLEKTLTGDGPAQIIAGFLGAPDVTSY 298
              LTMAPIAFVTMTEHPGHIMVLNSLTK+DYFK+PGLEKTLTGDG AQIIAGF+GAPDVTSY
    Sbjct: 245 LTMAPIAFVTMTEHPGHIMVLNSLTKKDYFKEPGLEKTLTGDGLAQIIAGFIGAPDVTSY 304

    Query: 299 GENIGVMALNKIFSVYVIAGAAVIAALLSFIGKVSALIQSIPTPVIGGISVALFGVIASS 358
              GENIGVMA+ KI S+YVIAGAAV+A ++SF+GK++AL+QSIP PVIGG S+ALFGVIA+S
    Sbjct: 305 GENIGVMAITKIHSIYVIAGAAVLAIVVSFVGKITALLQSIPAPVIGGASIALFGVIAAS 364

    Query: 359 GLKILIESKVDMDNKKNLLIASVILVSGIGGLMLQV-NGLQISGVAFSTLLGIILVQVLPE 417
              GLKIL+E+KVD D K+NLLI+SV+LV GIGG+++ +    LQIS VA +T+LGI+L  VLP+
    Sbjct: 365 GLKILVENKVDPDIKRNLLISSVVLVIGIGGMIINITQNLQISSVAIATTLGIVINLVLPK 425
```

An alignment of the GAS and GBS proteins is shown below.

```
     Identities = 186/425 (43%), Positives = 282/425 (65%), Gaps = 17/425 (4%)

Query: 30  NLLLDIDEKPELPQGLLLSPQHVFAMFGATILVPLILMMPVSVALFASGCGTLIVQVATK 89
           +++ D++E P+     + LSPQH+FAMFGAT+LVP+++G+  SVAL +SG GTL +   TK
Sbjct: 3   DVIYDVEEVPKAGMLVGLSPQHLFAMFGATVLVPILVGIDPSVALLSSGLGTLAHLSVTK 62

Query: 90  FKVPVYLGSSFAYITAMALAMKQMHGDISAAQTGILFVGLIYVVVATVIKFVGKSWVDKI 149
           FK+P Y+GSSFAYI AM L MK     I A   G +  GL+Y++VA ++K +GN W+D I
Sbjct: 63  FKIPAYMGSSFAYIAAMQLLMKT--NGIGAVAQGAMTGGLVYLIVALIVKAIGNDWIDNI 120

Query: 150 LPPIIIGPMIIVIGLGLANSAVTNAGFVAKGDWRK--MLVAVVTFLIAAFINTKHGFIK 207
           LPPI++GP+++VIGL LA++AV +   +  G++     +VT L   F N  GKG +
```

```
Sbjct: 121 LPPIVVGPIVMVIGLSLASTAVNDV-MLKNGNYNLTYLVIGLVTLLSVIFFNIYGKGIVA 179

Query: 208 IIPFLPAIGGYILSILLG------LVDLSFVEKAAWFELPKFYLPFKTGLFHSYKLYFG 261
           I+P L  ++ GY++++++G       +VD + V +A WF +P    +PF T     Y + F
Sbjct: 180 IVPLLLGLLVGVVVALLVGVLTSQEIVDFTNVAQAKWFSIPSVEIPFLT-----YGVKFV 234

Query: 262 PH-MLAILPISIVTIARNIGDHTVLGQICGRNFLKKPGLNRLLIGDGLATAFSALIGGFA 320
           P  +L + PI+ VT+ E+ G   VL  +  R++ K PGL + L GDG A    +G P
Sbjct: 235 PSAILTMAPIAFVTMTEHFGHIMVLNSLTKRDYFKDPGLEKTLTGDGFAQIIAGFLGAPP 294

Query: 321 ETTVGENTGVIGMTRIASVTVCRNAAFTAIAFSFFGKFTALISTIPSAVLGGMALLLYGV 380
            T+YGEN GV+ + +I SV VI  AA IA   SF GK +ALI +IP+ V+GG+++ L+GV
Sbjct: 295 VTSYGENIGVMALNKIFSVYVIAGAAVIAALLSFIGKVSALIQSIPTPVIGGISVALFGV 354

Query: 381 IASNGLKVLIENRVNFAEVRNLIIASSMLVLGLGGAVLDLGALTLSGTALSAIVGIILNL 440
           IAS+GLK+LIE+++V+     +NL+IAS +LV G+GG +L +  L +SG A S ++GIIL
Sbjct: 355 IASSGLKILIESKVDMNKKNLLIASVILVSGIGGLMLQVNGLQISGVAFSTLLGIILYQ 414

Query: 441 ILPKE 445
           +LP++
Sbjct: 415 VLPEK 419
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1749

A DNA sequence (GBSx1856) was identified in *S.agalactiae* <SEQ ID 5433> which encodes the amino acid sequence <SEQ ID 5434>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3863(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1750

A DNA sequence (GBSx1857) was identified in *S.agalactiae* <SEQ ID 5435> which encodes the amino acid sequence <SEQ ID 5436>. This protein is predicted to be sodium/alanine symporter. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.88   Transmembrane  191 - 207 ( 184 - 214)
   INTEGRAL    Likelihood = -8.97   Transmembrane  151 - 167 ( 148 - 171)
   INTEGRAL    Likelihood = -8.39   Transmembrane  217 - 233 ( 216 - 238)
   INTEGRAL    Likelihood = -6.74   Transmembrane  312 - 328 ( 310 - 333)
   INTEGRAL    Likelihood = -6.26   Transmembrane  357 - 373 ( 349 - 376)
   INTEGRAL    Likelihood = -5.10   Transmembrane  424 - 440 ( 422 - 441)
   INTEGRAL    Likelihood = -5.04   Transmembrane  396 - 412 ( 390 - 417)
   INTEGRAL    Likelihood = -0.37   Transmembrane   25 -  41 (  25 -  41)

----- Final Results -----
```

```
bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9867> which encodes amino acid sequence <SEQ ID 9868> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC22541 GB:U32770 amino acid carrier protein, putative
          [Haemophilus influenzae Rd]
 Identities = 255/443 (57%), Positives = 333/443 (74%), Gaps = 4/443 (0%)

Query: 11  TLPTHINSFVWGPPLLALLVGTGIYLSFRLGFIQLRQLSRAFKLIFREDNG-QGDISSYA 69
           ++ + I+SF+WG PLL LL GTG+YL+ RLGFIQ+R L RA   +F++D G +GD+SS+A
Sbjct: 5   SILSAIDSFIWGAPLLILLSGTGLYLTLRLGFIQIRYLPRALGYLFKKDKGGKGDVSSFA 64

Query: 70  ALATALAATVGTGNIVGVATAIKSGGPGALFWMWVAAPFGMATKYAEGLLAIKYRTKDTN 129
           AL TALAAT+GTGNIVGVATA++GGPGA+FWMW+ A  GMATKYAE LLA+KYR +D N
Sbjct: 65  ALCTALAATIGTGNIVGVATAVQAGGPGAIFWMWLVALLGNATKYAECLLAVKYRVRDKN 124

Query: 130 GEISGGPMYYIINGMGQKWKPLAVFFSAAGILVALLGIGTFTQVNAIASSLEHTFKISTR 189
           G ++GGPMYYI  G+G +W  LA  F+  G++VA  GIGTF QVNAI  +++ TF I
Sbjct: 125 GFMAGGPMYYIERGLGIRW--LAKLFALFGVMVAFFGIGTFPQVNAITHAMQDTFNIPVL 182

Query: 190 FTSLILAVIVLFIFGGIKSISKVSEKIVPFMAISYILATLIIIAVNYNKIPHTFQLIFS 249
            T+I+ ++V  II GG+K I+  S  IVPFMAI Y+  +L+II +N  K+P    LI
Sbjct: 183 VTAIVTLLVGLIILGGVKRIATASSVIVPFMAILYVTTSLVIILLNIEKVPDALLIID 242

Query: 250 GAFSGTAAIGGFSGAIVKEAIQKGIARGVFSNESGLGSAPIAAAAAKTKEPVEQGLISMT 309
           AF   AA+GG  G  V +AIQ G+ARG+FSNESGLGSAPIAAAAA+T+EPV QGLISMT
Sbjct: 243 SAFDPQAALGGAVGLTVMKAIQSGVARGIFSNESGLGSAPIAAAAAQTREPVRQGLISMT 302

Query: 310 GTFIDTIVICTLTGIAILVTGKWLEFDLQGAPLTQASFNTVFG-SLGSPALTFCLVLFAF 368
           GTF+DTI++CT+TGI +++TG W   +L GA +T  +F    G S+G+  +T  L+ FAF
Sbjct: 302 GTFLCTIIVCTMTGIVLVLIGAWNPELAGATVTNYAFAQGLGTSIGATITVTVGLLFFAF 362

Query: 369 TTILGWSYYGERCFEYLFGTKFINAYRIIPVIMVGLGGFLQLDLIWVIADIVNGLMALPN 428
           TTILGW YYGERCP YL G + +  YR+ ++++VGLG FL L+LIW+IADIVNGLMA PN
Sbjct: 363 TTILGWCYYGERCFVYLVGIRGVKLYRLAYIMLVGLGAFLHLNLIWIIADIVNGLMAFPN 422

Query: 429 LIALLALSPIIVKETQKYFSETK 451
           LIAL+ L  +I++ET+ YF   K
Sbjct: 423 LIALIGLRKVIIEETKDYFQRLK 445
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5437> which encodes the amino acid sequence <SEQ ID 5438>. Analysis of this protein sequence reveals the following:

```
      Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood =-11.36   Transmembrane  183 - 199 ( 175 - 206)
   INTEGRAL     Likelihood = -7.80   Transmembrane  143 - 159 ( 140 - 163)
   INTEGRAL     Likelihood = -7.11   Transmembrane  209 - 225 ( 208 - 229)
   INTEGRAL     Likelihood = -5.95   Transmembrane  416 - 432 ( 413 - 434)
   INTEGRAL     Likelihood = -5.15   Transmembrane  304 - 320 ( 302 - 324)
   INTEGRAL     Likelihood = -4.46   Transmembrane  387 - 403 ( 382 - 408)
   INTEGRAL     Likelihood = -3.35   Transmembrane  348 - 364 ( 345 - 366)
   INTEGRAL     Likelihood = -1.17   Transmembrane   11 -  27 (  10 -  28)

----- Final Results -----
              bacterial membrane --- Certainty=0.5543(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF94579 GB:AEC04221 sodium/alanine symporter [Vibrio cholerae]
 Identities = 261/441 (59%), Positives = 328/441 (74%), Gaps = 7/441 (1%)

Query: 3   ALVKLIDNLVWGPPLLILLVGTGIYLTSHLGLIQILKLPRAFKLIFSDDEG---HGDISS 59
            + ++ +D+LVWGPPLLILLVGTG+Y T  LGL+Q  +LP A  ++F ++      GD+SS
Sbjct: 6   SFLQTVDSLVWGPPLLILLVGTGVYFTFRLGLLQFRRLPTALAMVPGREKSSDKQGDVSS 65

Query: 60  FAALATALAATVGTGNIVGVATAIKSGGPGALPWMWVAAFPGMATKYAEGVLAIKVRTKD 119
            FAAL TAL+AT+GTGNIVGVATAIK GGPGALPWM+AA PGMATKYAE +LA+KVR  D
Sbjct: 66  FAALCTALSATIGTGNIVGVATAIKLGGPGALPWMMLAALPGMATKYAECLLAVKYRQID 125

Query: 120 ANGHISGGSMYYIVNGMGTKWKPLAVLFAGSGILVALPGIGTFAQVNSITSSLGHSPGLS 179
             G + GGPMYY+ +G+ +K   LAVLFA  + VA PGIGTF QVN+I +  SPG+
Sbjct: 126 DKGSMVGGPMYYLRDGVSSK--TLAVLFAVPAVGVACPGIGTFPQVNAILDATQISPGVP 183

Query: 180 PQMVSIVLAIFVAAIIPGGIHSISKVAEKVVPFMAIFYILSSLAVIPSHYQQLLEPVIRLV 239
              +  ++VL + VA +  GGI SI+KVA KVVP MA+FYI++ L+VI ++  +L   + LV
Sbjct: 184 REASAVVLTVLVAIVTIGGIQSIAKVAGKVVPAMALFYIACLSVIVTNADKLADAVELV 243

Query: 240 FQSAFTPTAAIGGFAGSLMKDAIQKGIARGVFSNESGLRSAPIAAAAAKTNEPVEQGLIS 299
              SAFT TAA GGF G+ +  AIQ GIARGVFSNESGL SAP+AAAAAKT+  VEQGLIS
Sbjct: 244 LVSAFTSTAATGGFLGASIMLAIQSGIARGVFSNESGLSSAPMAAAAAKTDSCVEQGLIS 303

Query: 300 MIGTFIDTIICTLTGLSILVTGQWTGQLEGAPLTQSAFATVVG--NLGTPGLTFSLVLF 357
            MIGTF DTIICT+TGL++++TG W   L GA +T  AFAT      +G   ++  L+ F
Sbjct: 304 MIGTFFDTIICTWTGLALLLTGAWQSDLSGAAMTTYAFATGLNAQTIGPMLVSIGLMFF 363

Query: 358 AFTTILGWSYYGERCPEFLPGITHLTYFRIVFILMVGLGGFLKLELIWVLADIVNGLMAL 417
            AFTTILGW+YYGERC  FLPG   + ++IVFI ++  G FL L+LIW++ADIVNGLMA+
Sbjct: 364 AFTTILGWMYYGERCMVFLPGTKAVLPYKIVFIGLIASGAFLHLDLIWIIADIVNGLMAI 423

Query: 418 PNLIALLALSPVVILETKHYF 438
            PNLI L+AL  VV+ ETK YF
Sbjct: 424 PNLIGLVALRHVVVEETKQYF 444
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 323/439 (73%), Positives = 380/439 (85%), Gaps = 1/439 (0%)

Query: 9   MLTLFTHINSFVWGPPLLALLVGTGIYLSPRLGHIQLRQLSRAFKLIFREDNGQGDISSY 68
            M+ L   I++ VWGPPLL LLVGTGIYL+  LG IQ+ +L RAFKLIF +D G GDISS+
Sbjct: 1   MIALVKLIDNLVWGPPLLILLVGTGIYLTSHLGLIQILKLPRAFKLIFSDDEGHGDISSF 60

Query: 69  AALATALAATVGTGNIVGVATAIKSGGPGALPWMWVAAFPGMATKYAEGLLAIKYRTKDT 128
            AALATALAATVGTGNIVGVATAIKSGGPGALPWMWVAAFPGMATKYAEG+LAIKYRTKD
Sbjct: 61  AALATALAATVGTGNIVGVATAIKSGGPGALPWMWVAAFPGMATKYAEGVLAIKYRTKDA 120

Query: 129 NGEISGGPMYYIINGMGQKWKPLAVFPSAAGILVALLGIGTFQVNAIASSLEHTFKIST 188
            NG ISGGPMYYI+NGMG KWKPLAV F+ +GILVAL GIGTF QVN+I SSL H+F +S
Sbjct: 121 NGHISGGPMYYIVNGMGTKWKPLAVLFAGSGILVALPGIGTFAQVNSITSSLGHSFGLSP 180

Query: 189 RFTSLIIAVIVLFIIFGGIKSISKVSEKIVPFMAISYILATLIIIAVNYNKIPHTFPQLIF 248
            + S++LA+ V  IIFGGI SISKV+EK+VPFMAI YIL++L +I  +Y ++     +L+F
Sbjct: 181 QMVSIVLAIFVAAIIPGGIHSISKVAEKVVPFMAIFYILSSLAVIFSHYQQLLPVIRLVF 240

Query: 249 SGAPSGTAAIGGFSGAIVKEAIQKGIARGVFSNESGLGSAPIAAAAAKTKEPVEQGLISM 308
             AF+  TAAIGGF+G+++K+AIQKGIARGVFSNESGL SAPIAAAAAKT EPVRQGLISM
Sbjct: 241 QSAFTPTAAIGGFAGSLMKDAIQKGIARGVFSNESGLRSAPIAAAAAKTNEPVEQGLISM 300

Query: 309 TGTFIDTIVICTLTGIAILVTGKWLEFDLQGAPLTQASFNTVPGSLGSFALTFCLVLFAF 368
            TGTFIDTI+ICTLTG++ILVTG+W    L+GAPLTQ++F TVPG+LG+F LTF LVLFAF
Sbjct: 301 TGTFIDTIICTLTGLSILVTGQWTG-QLEGAPLTQSAFATVPGNLGTFGLTFSLVLFAF 359

Query: 369 TTILGWSYYGERCPEYLPGTKPINAYRIIFVIMVGLGGFLQLDLIWVIADIVNGLMALPN 428
            TTILGWSYYGERCPE+LPG   + +RI+F++MVGLGGFL+L+LIWV+ADIVNGLMALPN
Sbjct: 360 TTILGWSYYGERCPEFLPGITHLTYFRIVFILMVGLGGFLKLELIWVLADIVNGLMALPN 419

Query: 429 LIALLALSPIIVKETQKYF 447
```

```
            LIALLALSP+++ ET+ YF
Sbjct: 420 LIALLALSPVVILETKHYF 438
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1751

A DNA sequence (GBSx1858) was identified in *S.agalactiae* <SEQ ID 5439> which encodes the amino acid sequence <SEQ ID 5440>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.16   Transmembrane   85 - 101 (  80 - 108)
   INTEGRAL    Likelihood = -5.36   Transmembrane  118 - 134 ( 115 - 137)
   INTEGRAL    Likelihood = -2.81   Transmembrane  177 - 193 ( 177 - 193)
   INTEGRAL    Likelihood = -0.48   Transmembrane   49 -  65 (  49 -  65)

----- Final Results -----
              bacterial membrane --- Certainty=0.3463(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12451 GB:Z99107 alternate gene name: ydxT-similar to caticn
          efflux system membrane protein [Bacillus subtilis]
 Identities = 118/282 (41%), Positives = 181/282 (63%)

Query: 6   ENLQLAKRGPIISIIAYITLAVAKLAAGYWFDATSLVADGFNNLSDILGNVALLIGLHLA 65
           + L+  + G ++SI AY+ L+  KL  GY F + +L ADG NN +DI+ +VA+LIGL ++
Sbjct: 5   DELKKGESGALVSIAAYLVLSAIKLIIGYLFHSEALTADGLNNVTTDIIASVAVLIGLRIS 64

Query: 66  SQPADSNNRPGHWKIEDLASLITSFIMFVVGIQVPIQTVTKIINNTDTNIDDLGAIVGAI 125
            +P D +H +GH++ E +ASLI SFIM VVG+QV       I +      D + A   A
Sbjct: 65  QKPPDEDHPYGHFRAETIASLIASFIMVVGLQVLFSAGESIFSAKQETPDMIAWTAAG 124

Query: 126 SALVMLGVYFYNKQLSQRVKSSALVAASKDNLSDAVTSIGTSIAIIAASLNFPIIDRLAA 185
            A++ML VY YNK+L+++VKS AL+AA+ DN SDA  SIGT I I+AA  +   ID + A
Sbjct: 125 GAVLMLIVYRYNKRLAKKVKSQALLAAAADNKSDAFVSIGTFIGIVAAQFHLAWIDTVTA 184

Query: 186 IIITYFILKTAYDIFIESAFSLSDGFDDYQLKQYEKAILTIPKISAVKSQRGRTYGSNIY 245
            +I   I KTA+DIF ES+ SL+DGFD   +  Y++ I  I  +S +K  + R  GS ++
Sbjct: 185 FVIGLLICKTAWDIFKESSHSLTDGFDIKDISAYKQTIEKISGVSRLKDIKAYLGSTVH 244

Query: 246 LDIVLEMNPDLSVFESHAITERVEKLLSDKFSVYDIDIHVEP 287
           +D+V+E++ DL++ ESH I   +E+ + ++ ++     +H+EP
Sbjct: 245 VDVVVEVSADLNITESHDIANEIERRMKEEHAIDYSHVIMEP 286
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5441> which encodes the amino acid sequence <SEQ ID 5442>. Analysis of this protein sequence reveals the following:

```
      Possible site: 46
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.01   Transmembrane  121 - 137 ( 114 - 139)
   INTEGRAL    Likelihood = -5.41   Transmembrane   86 - 102 (  84 - 109)
   INTEGRAL    Likelihood = -5.04   Transmembrane  178 - 194 ( 176 - 197)
   INTEGRAL    Likelihood = -0.69   Transmembrane   50 -  66 (  50 -  66)
   INTEGRAL    Likelihood = -0.64   Transmembrane  158 - 174 ( 158 - 174)

----- Final Results -----
              bacterial membrane --- Certainty=0.4206(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12451 GB:Z99107 alternate gene name: ydxT-similar to cation
          efflux system membrane protein [Bacillus subtilis]
 Identities = 127/280 (45%), Positives = 187/280 (66%)

Query: 9   LKLARKGPIVSIIVYLSLSVAKLLAGYLLNASSLIADGFNMLSDIVGNVALLIGLHLASQ 68
            LK    G +VSI  YL LS  KL+ GYL ++ +L ADG NN +DI+ +VA+LIGL ++ +
Sbjct: 7   LKKGRSGALVSIAAYLVLSAIKLIIGYLPHSSALTADGLNNFTDIISAVAVLIGLRISQK 66

Query: 69  PADANHKFGHWKIEDLSSLVTSFIMFLVGFQVLIHTIKSIPSGQQVDIDPLGAIVGIVSA 128
            P D +H +GH++ E ++SL+ SFIM +VG QVL     +SIFS +Q   D + A      A
Sbjct: 67  PPDEDHPYGHPRAETIASLIASFINMVVGLQVLFSNGESIFSAKQETPDMIAAWTAAGSA 126

Query: 129 FVNLGVYVFNKRLSKRVKSSALVAASKDNLADAVTSIGTSIAIIAASLHLPVIDHIAAMI 188
             +ML VY +NKRL+K+VKS AL+AA+ DN +DA  SIGT I I+AA  HL  ID + A +
Sbjct: 127 VLMLVYRYNKRLAKKVKSQALLAAAADNKSDAFVSIGTFIGIVAAQFHLAWIDTVTAFV 186

Query: 189 ITFFILKTAFDIFMESSFSLSDGFDSRHLKKYRKAILEIPKIVAVKSQKAKTYGSNVYLD 248
            I   I KTA+DIF ESS SL+DGFD + +  Y++ I +I  +  +K  +AR  GS V++D
Sbjct: 187 IGLLICKTAWDIFKESSHSLTDGFDIKDISAYKQTIEKISGVSRLKDIKARYLGSTVHVD 246

Query: 249 IVLENNPDLSVYESHSITTEKVEQLLESDQPSIYDIDIHVEP 288
            +V+E++ DL++ ESH I  ++E+ + ++ +I    +H+EP
Sbjct: 247 VVVEVSADLNITESHDIANEIERRMKEEHAIDYSHVEMEP 286
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 274/406 (67%), Positives = 340/406 (83%), Gaps = 4/406 (0%)

Query: 7   NLQLAKRGPIISIIAYITLAVAKLAAGYWFDATSLVADGFNNLSDILGNVALLIGLHLAS 66
            NL+LA++GPI+SII Y++L+VAKL AGY  +A+SL+ADGFNNLSDI+GNVALLIGLHLAS
Sbjct: 8   NLKLARKGPIVSIIVYLSLSVAKLLAGYLLNASSLIADGFNNLSDIVGNVALLIGLHLAS 67

Query: 67  QPADSNHRFGHWKIEDLASLITSFIMFVVGIQVFIQTVTKIINNTDTNIDPLGAIVGAIS 126
            QPAD+NH+FGHWKIEDLASL+TSFIMF+VG QV I T+  I +   +IDPLGAIVG +S
Sbjct: 68  QPADANHKFGHWKIEDLSSLVTSFIMFLVGFQVLIHTIKSIPSGQQVDIDPLGAIVGIVS 127

Query: 127 ALVMLGVYFVNKQLSQRVKSSALVAASKDNLSDAVTSIGTSIAIIAASLF+ P+ID +AA+
            A VMLGVY +NK+LS+RVKSSALVAASKDNL+DAVTSIGTSIAIIAASL+ P+ID +AA+
Sbjct: 128 AFVMLGVYVFNKRLSKRVKSSALVAASKDNLADAVTSIGTSIAIIAASLHLPVIDHIAAM 187

Query: 187 ITTYFILKTAYDIFIESSAFSLSDGFDDYQLKQYEKAILTIPKISAVKSQRGRTYGSNIYL 246
            ITT+FILKTA+DIF+ES+FSLSDGFD   LK+YEKAILTIPKI AVKSQR RTYGSN+YL
Sbjct: 188 ITTFFILKTAFDIFMESSFSLSDGFDSRHLKKYEKAILEIPKIVAVKSQRARTYGSNVYL 247

Query: 247 DIVLEMNPDLSVFESHAITERVEKLLSDKPSVYDIDIHVEPASIPEDEIFDNVYQKLYKN 306
            DIVLEMNPDLSV+ESH+ITE+VE+LLSD+PS+YDIDIHVEPA IPE+EIFDNV +KLY+
Sbjct: 248 DIVLEMNPDLSVYESHSITEKVEQLLSDQPSIYDIDIHVEPAMIPEEEIFDNVVAKKLYRY 307

Query: 307 EKIILAKIPGYETFISPDFYMINKKGNIITSDWLTNATNHSLASNFKYPNVKSISQKTKL 366
            EK+ILAKIPGYETFISPDFYMINKKGNIITSDWLTNATNHSLASNFKYPNVKSISQKTKL
Sbjct: 308 EKLILSKVPDYDHYIAKSFQLIDANGQTVNYEQFLNQSIY-YPGNFNHFQIESISQKTML 366

Query: 367 VSYELEGKRHTSIWRRNEKWFLIYHQIT--AKSSPYKTRRYQITSL 410
            V+Y+L G + TSIWRR+E W L++HQIT  AK   + T  Y+I  +
Sbjct: 367 VTYQLNGNQRTSIWRRHESWSLLFHQITPIAKKQLHHT-HYRIVKM 411
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1752

A DNA sequence (GBSx1859) was identified in *S.agalactiae* <SEQ ID 5443> which encodes the amino acid sequence <SEQ ID 5444>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.12   Transmembrane  171 - 187 ( 161 - 194)
   INTEGRAL    Likelihood = -6.32   Transmembrane  118 - 134 ( 113 - 138)
   INTEGRAL    Likelihood = -5.89   Transmembrane   59 -  75 (  53 -  77)
   INTEGRAL    Likelihood = -5.52   Transmembrane  231 - 247 ( 226 - 252)
   INTEGRAL    Likelihood = -3.24   Transmembrane   86 - 102 (  84 - 103)
   INTEGRAL    Likelihood = -0.32   Transmembrane   31 -  47 (  31 -  47)

----- Final Results -----
               bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9869> which encodes amino acid sequence <SEQ ID 9870> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14850 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 80/226 (35%), Positives = 136/226 (59%), Gaps = 1/226 (0%)

Query: 27  TNNPIFGIMLTVWAYYIGIRIFRKYPSPAT-TPLLLATILLIAFLKL/THISYKDYYNGGS 85
           T +P FGI++++ A+ IG  +F+K       TPL +A +L IAFLK+   SY DY NGG
Sbjct: 4   TMSPYFGIVVSLAAFGIGTFLFKKTKGFFLFTPLFVAMVLGIAFLKIGGFSYADYNNGGE 63

Query: 86  FLTMLITPSTVVLAIPLYRTFHLMKHHIKSISISIILASVINTVFTAIVAKFFGMKYFLA 145
            +   + P+T+  AIPLY+    +K +   I  SII  S+ +     ++AK   +   +
Sbjct: 64  IIKFFLEPATIAFAIPLYKQRDKLKKYWWQIMASIIAGSICSVTIVYLLAKGIHLDSAVM 123

Query: 146 ISLFPKSVTTAMAVGITSKAGGLATITLVVVVITGILTSVLGPIFLKLLRIEDPVAIGLA 205
            S+ P++ TTA+A+ ++   GG++ IT   V+   ++    LG +FLK+ ++++P++ GLA
Sbjct: 124 KSNLPQAATTAIALPLSKGIGSISDITAFAVIFNAVIVIALGALFLKVFKVKNPISKGLA 183

Query: 206 LGGTGHAIGTGQALKYGQVGAMAGLAIGITGICVVISPLVAGLI 251
           LG +GHA+G    ++ G+V+ AMA +A+ + G+  V+V P+   LI
Sbjct: 184 LGTSGHALGVAVGIEMGEVEAMASIAVVVVGVVTVLVIPVFVQLI 229
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8893> and protein <SEQ ID 8894> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
SRCFLG: 0
McG: Length  of  UR:   22
     Peak Value of UR:   2.57
     Net Charge of CR: 0
McG: Discrim Score:      6.51
GvH: Signal Score (-7.5): -5.91
     Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 6 value:  -8.12 threshold:  0.0
   INTEGRAL    Likelihood = -8.12   Transmembrane  149 - 165 ( 139 - 172)
   INTEGRAL    Likelihood = -6.32   Transmembrane   96 - 112 (  91 - 116)
   INTEGRAL    Likelihood = -5.89   Transmembrane   37 -  53 (  31 -  55)
   INTEGRAL    Likelihood = -5.52   Transmembrane  209 - 225 ( 204 - 230)
   INTEGRAL    Likelihood = -3.24   Transmembrane   64 -  80 (  62 -  81)
   INTEGRAL    Likelihood = -0.32   Transmembrane    9 -  25 (   9 -  25)
   PERIPHERAL  Likelihood =  1.06     121
 modified ALOM score:   2.12
icm1 HYPID: 7  CFP: 0.425

*** Reasoning Step: 3
```

```
----- Final Results -----
                bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01066(325 - 999 of 1305)
EGAD|107753|BS2884(4 - 229 of 231) hypothetical protein (Bacillus subtilis) OMNI|NT01BS3363
LrgB    GP|1770004|emb|CAA99613.1||Z75208   hypothetical   protein   (Bacillus   subtilis)
GP|2635355|emb|CAB14850.1||Z99118 similar to hypothetical proteins (Bacillus subtilis)
PIR|D69983|D69983 conserved hypothetical protein ysbB - Bacillus subtilis
%Match = 17.2
%Identity = 35.4  %Similarity = 62.4
Matches = 80  Mismatches = 84  Conservative Sub.s = 61

192       222       252       282       312       342       372       402
WSTFKT*SPIFLG*LSLS*ERYFSIF*LLDWYPWGSKRDMKEIIQKLEVKMATLTNNPIFGIMLTWVAYYIGIRIFRKYP
                                                                 | :| |||:::: |: ||  :|:|
                                                               MESTMSPYFGIVVSLAAFGIGTFLFKKTK
                                                                         10        20

429       459       489       519       549       579       609       639
SPAT-TPLLLATILLIAFLKLTHISYKDYYNGGSFLTMLATPSTVVLAIFLYRTFHIMKKHHIKSISISIILASVINTVFT
      |||:|| :| ||||||:   || || ||| :  :: |:|: :|||||:   :|  :   |   |||  |: :
GFFLFTPLFVAMVLGIAFLKCGGFSVADYNKGGEIIKFLEPATIAFAIPLYKQRDELKKYWWQIASIIAGSICSVTIV
          40        50        60        70        80        90       100

669       699       729       759       789       819       849       879
AIVAKFFGMKYFLAISLFPKSVTTAMAVGITSKAGGLATITLVVVVITGILTSVLGPIFLKLLRIEDPVAIGLALGGTGH
 ::||   :   :  |::|:: |||:|: ::   ||:: ||   |:   ::   || :|||::::::|:: ||||| :||
YLLAKGIHLDSAVMKSMLPQAATTALALPLSKGIGGISDITAFAVIFNAVIVVALGALFLKVFEVKNPISKGLALGTSGH
         120       130       140       150       160       170       180

909       939       969       999       1029      1059      1089      1119
AIGTGQALKYGQVCGAMAGLAIGITGICVVIVSPLVAGLILK*G*GK*TQNNYVIIFKNRI*DK*L*YR*KK*LLERLSV
|:|   :: |:|: ||| :|: :  |:|  |:  ||
ALGVAVGIEMGEVERAMASIAVVVVGVVTVLVIPVFVQLIGG
         200       210       220       230
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1753

A DNA sequence (GBSx1860) was identified in *S.agalactiae* <SEQ ID 5445> which encodes the amino acid sequence <SEQ ID 5446>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> May be a lipoprotein

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>gp:CAA76857 GB:Y17797 hypothetical protein [Enterococcus faecalis]
 Identities = 44/194 (22%), Positives = 90/194 (45%), Gaps = 13/194 (6%)

Query: 21  TACSSSNTQQTSTSKSNVSQHKNIKADHERLRLKPNKVKLGVKANNPKGGTSLAELKQLF 80
           T  S ++T++ S+ K +  + K    D+ +L+  ++K+ +G   N+ +GG++  E+K +
Sbjct: 60  TNSSKNDTKKESSEKKSRDKSK----DNSDLKATYDKINVGDIMNSSEGGSTRDEVKAIL 115

Query: 81  GGEPNEKFDTPAGNVTLKGYRW-NVDD----ISITIQLLNDSSIVRSISNFKPIRDANIT 135
```

```
              GEP     T     ++      W NV       SIT+   +  +  +S+S  K  +    +T
Sbjct: 116 -GEPASSSTTDIQGISTTTLSWTNVKGGDLLASITVSFSDGKAASKSVSGLKVAKHDKVT 174

Query: 136 TEDYNSLKNGMSYN--KVKELLGEPDDISQAVSSDKEELQAAWISGIQSSDSDPGINLTF 193
            N+=    SY+  + ++ LG+P  I+     + ++      W+  +   D    + ++F
Sbjct: 175 ADQVNNIATDGSYSEEQARKDLGDPTGITSTNINGEKNDTLIWMKNL-DGDLGATVTVSF 233

Query: 194 ENDKLTNKQQHGLK 207
             N    +K   GLK
Sbjct: 234 SNGNAISKSSSGLK 247
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5447> which encodes the amino acid sequence <SEQ ID 5448>. Analysis of this protein sequence reveals the following:

```
      Possible site: 21

>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA76857 GB:Y17797 hypothetical protein [Enterococcus faecalis]
 Identities = 34/166 (20%), Positives = 74/166 (44%), Gaps = 8/166 (4%)

Query: 47  HQDKRANFEKIKLATVDSSFTGGTSLEELISLFGEPSQHDPKTAGEVTIDAYTWQFDQ-- 104
           + D +A ++KI +   + +S  GG++ +E+ ++ GEP+        ++      +W    +
Sbjct: 83  NSDLKATYDKINVGDIMNSSEGGSTEDEVKQLLGEPASSSTTDIQGISTTTLSWTNVKGG 142

Query: 105 ---VTLTVNLYQNSSIVKTISNFTFARELGLSQKEYQQLQKGMSY--EDVKKILTEPDNY 159
               ++TV+     +  K++S    A+   ++  +    +   SY  E  +K L +P
Sbjct: 143 DLLASITVSFSDGKAASKSVSGLKVAKHDKVTADQVNNIATDGSYSEEQARKDLGDPTGI 202

Query: 160 SQASSSDHQTLQAIWVSGLKTDTSGANISLVFENNQLTEMSQVGLE 205
           +  + +  +    IW+  L  D  GA +++ F N       S  GL+
Sbjct: 203 TSTNINGEKNDTLIWMKNLDGDL-GATVTVSFSNGNAISKSSSGLK 247
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 84/199 (42%), Positives = 126/199 (63%), Gaps = 3/199 (1%)

Query: 11  TIVCLSFLG--LTACSSSNTQCTSTSKSNVSQHKNIKADHEELRLKFNKVKLGVKANNFK 68
           T++ +SF    L ACS++  ++   S S +   +  +A H++ R  F K+KL     +F
Sbjct: 8   TLLLISFFTSFLVACSTTKDKEPQPSDSEIITPRLHQAAHQDKRANFEKIKLATVDSSFT 67

Query: 69  GGTSLAELKQLFGGEPNEKPDTPAGNVTLKGYRWNVDDISITIQLLNDSSIVRSISNFKF 128
           GGTSL EL  LFG EP++      AG VT+  Y W  D +++T+ L  +SSIV++ISNF F
Sbjct: 68  GGTSLEELISLFG-EPSQHDPKTAGEVTIDAYTWQFDQVTLTVNLYQNSSIVKTISNFTF 126

Query: 129 IRDANITTKDYNSLKNGMSYNKVKELLGEPDDISQAVSSDKEELQAAWISGIQSSDSDPG 188
            R+  ++ K+Y  L+ GMSY  VK++L EPD+ SQA SSD + LQA W+SG+++  S
Sbjct: 127 ARELGLSQKEYQQLQKGMSYEDVKKILTEPDNYSQASSSDHQTLQAIWVSGLKTDTSGAN 186

Query: 189 INLTFENDKLTNKQQHGLK 207
           I+L FEN++LT   Q GL+
Sbjct: 187 ISLVFENNQLTEMSQVGLE 205
```

SEQ ID 5446 (GBS650) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 178 (lane 9; MW 28kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1754

A DNA sequence (GBSx1861) was identified in *S.agalactiae* <SEQ ID 5449> which encodes the amino acid sequence <SEQ ID 5450>. This protein is predicted to be ribosomal protein S1 homolog; Sequence specific DNA-binding protein (r. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2950(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9363> which encodes amino acid sequence <SEQ ID 9364> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA97575 GB:U27517 ribosomal S1 protein [Homo sapiens]
 Identities = 156/305 (51%), Positives = 214/305 (70%), Gaps = 7/305 (2%)

Query: 1   MEARKAWDKLWGRBGEVVTVKGTRAVKGGLSVEPBGLRGFIPASMIDTRPVRNTEKFVGQ 60
           ++ARKAW+ L   EG+ V K   AV+GGL V+  G+RGF+PASM+  RPV +  +P  +
Sbjct: 53  LDARKAWENLSFABGDTVDAKVINAVRGGLIVDVNGVRGFVPASMVAERFVSDLNQFKNK 112

Query: 61  EFDAKIKEVDAAENRPILSRREVVEESAAAARKEVFSNIEVGSVVTGKVARLTSFGAFID 120
           +  A++ E+D A  R ILSR+ V  +  AA   EVFS + VG VV G VARLT FGAF+D
Sbjct: 113 DIKAQVIEIDPANARLILSRKAVAAQERAAQLAEVFSKLSVGEVVEGTVARLTDFGAFVD 172

Query: 121 LGGVDGLVHVTELSHERNVSPKSVVTVGEEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV 180
           LGGVDGLVHV+E+SH+R  +P  V+T G++V+VK+L++D E GR+SLS+KAT  GPWD
Sbjct: 173 LGGVDGLVHVSEISHDRVKNPADVLTKGDKVDVKILALDTEKGRISLSIKATQRGPWDEA 232

Query: 181 EQKLAAGDVIEGKVKKLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVLSAGQEVTVK 240
            ++AAG V+EG VKR+ DFGAFVE+LPGI+GLVH+SQISK+R+ENP +VL +G +V VK
Sbjct: 233 ADQIAAGSVLEGTVKRVKDFGAFVEILPGIRGLVHVSQISNKRIENPSEVLKSGDKVQVK 292

Query: 241 VLEVNSDAERVSLSMKALEEKPAQAEGEKEEKKRQSRPRRPRKQKKRDYELPETQTGFSMA 300
           VL++     ER+SLSMKALEE+P       + E R+       R +   Y+  +  +  ++
Sbjct: 293 VLDIKPAEERISLSMKALEEKP------EREDRRGNDGSASRADIAAYK-QQDDSAATLG 345

Query: 301 DLFGD 305
           D+FGD
Sbjct: 346 DIFGD 350
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5451> which encodes the amino acid sequence <SEQ ID 5452>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3312(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 284/309 (91%), Positives = 296/309 (94%), Gaps = 1/309 (0%)

Query: 1   MEARKAWDKLVGRRGEVVTVKGTRAVKGGLSVEPEGLRGFIPASMIDTRPVRNTEKPVGQ 60
           +EARKAWDKLVGRRGEVVTVKGTRAVKGGLSVEPEGLRGFIPASMIDTRPVRNTEKPVGQ
Sbjct: 93  LEARKAWDKLVGRRGEVVTVKGTRAVKGGLSVEPEGLRGFIPASMIDTRPVRNTEKPVGQ 152
```

```
Query: 61  EFDAKIKEVDAAENRPILSRREVVEESAAAARKEVFSNIEVGSVVTGKVARLTSFGAFID 120
           EFDAKIKEVDAAENRPILSRREV+EE+A  AR EVFS I  G+VVTG VARLTSFGAFID
Sbjct: 153 EFDAKIKEVDAAENRPILSRREVIEEAAKSARAEVFSKISRGAVVTGTVARLTSFGAFID 212

Query: 121 LGGVDGLVHVTELSHERNVSPKSVVTVGEEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV 180
           LGGVDGLVHVTELSHERNVSPKSVV+VGKEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV
Sbjct: 213 LGGVDGLVHVTELSHERNVSPKSVVSVGKEVEVKVLSIDEEAGRVSLSLKATTPGPWDGV 272

Query: 181 EQKLAAGDVIEGKVKRLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVLSAGQEVTVK 240
           EQKLA GDV+EGKVKRLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVLS GQEVTVK
Sbjct: 273 EQKLAQGDVVEGKVKRLTDFGAFVEVLPGIDGLVHISQISHKRVENPKDVLSVGQEVTVK 332

Query: 241 VLEVNSDAERVSLSMKALEERPAQAEGE-KEEKRQSRPRRPRRQKKRDYELPETQTGFSM 299
           VLEVN+  ERVSLS+KALEERPAQAEG+ KEEKRQSRPRRP+R+ +RDYELPETQTGFSM
Sbjct: 333 VLEVNAADERVSLSIKALEERPAQAEGDNKEEKRQSRPRRPKRRSRRDYELPETQTGFSM 392

Query: 300 ADLFGDIEL 308
           ADLFGDIEL
Sbjct: 393 ADLFGDIEL 401
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1755

A DNA sequence (GBSx1862) was identified in *S.agalactiae* <SEQ ID 5453> which encodes the amino acid sequence <SEQ ID 5454>. This protein is predicted to be dihydroorotate dehydrogenase a (pyrD). Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
        bacterial cytoplasm --- Certainty=0.1708(Affirmative) < succ>
         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB51330 GB:AJ131985 dihydroorotate dehydrogenase [Streptococcus pneumoniae]
 Identities = 227/310 (73%), Positives = 268/310 (86%)

Query: 1   MVSLKTEIAGPSPDNCLMNAAGIYCMTKEELLAIENSEAGSFVTKTGTLEAREGNPQPRY 60
           MVS KT+IAGP PDNCLMNAAG+ CMT EEL  ++NS AG+FVTKT TL+ R+GNP+PRY
Sbjct: 1   MVSTKTQIAGPEPDNCLMNAAGVACMTIEELEEVKNSAAGTFVTKTATLDPRQGNPEPRY 60

Query: 61  ADTDWGSINSMGLPNKGIDYYLDFVTELQDQNSKNHVLSLVGLSPEETHIILKKVENSS 120
            D   GSINSMGLPN G+DYYLD++ +LQ++++++    LSLVG+SPEETH ILKKV+ S
Sbjct: 61  QDVPLGSINSMGLPNNGLDYYLDYLLDLQEKESMRTPFLSLVGMSPEETHTILKKVQESD 120

Query: 121 YNGLIELNLSCPNVPGKPQIAYDFSMTDLILSRIFSYYQKPLGIKLPPYFDIVHFDQAAT 180
           + GL ELNLSCPNVPGKPQIAYDFS TD IL+E+F+Y+ KPLGIKLPPYFDIV+FDQAA
Sbjct: 121 FRGLDTKLNLSCPNVPGKPQIAYDFSTTDRILAEVFAYFTKPLGIKLPPYFDIVYFDQAAA 180

Query: 181 IFNKYPLAFINCVNSIGNGLVIDDETVVIKPKNGFGGIGGDPIKPTALANVHAFVKRLNP 240
           IFNKYPL F+NCVNSIGNGL I+DE+VVI+PKNGFGGIGG++IKPTALANVHAFV+RLNP
Sbjct: 181 IFNKYPLKFVNCVNSIGNGLYIKDESVVIRPKNGFGGIGGKYIKPTALANVHAFVQRLNP 240

Query: 241 SIKIIGTGGVKNGRDAFEHILCGASMVQIGTALQKEGPEIFQRVSRELKEIMADKGYQSL 300
            I+IIGTGGV  GRDAFEHILCGASMVQ+GT L KEG   F R++  ELK IM +KGY+SL
Sbjct: 241 QIQIIGTGGVLTGRDAFEHILCGASMVQVGTTLHKEGVSAFDRITNELKAIMVEKGYRSL 300

Query: 301 EDFRGQLNYL 310
           EDFRG+L Y+
Sbjct: 301 EDFRGKLRYI 310
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5455> which encodes the amino acid sequence <SEQ ID 5456>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2689(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 239/309 (77%), Positives = 262/309 (84%)

Query: 1   MVSLKTEIAGPSFDNCLMNAAGIYCMTKEELLAIENSKAGSFVTKTGTLKAEGNPQPRY 60
            MVS  T+I  FSFDNCLMNAAG+YCMTKEEL+ +E S+A SFVTKTGTLE R GNP+PRY
Sbjct: 5   MVSTATQIGHFSFDNCLMNAAGVYCMTKEELMSVEKSQAASFVTKTGTLEVRPGNPEPRY 64

Query: 61  ADTDWGSINSMGLPNKGIDYYLDFVTELQDQDNSKNHVLSLVGLSPEETHIILKKVENSS 120
            ADT  GSINSMGLPN G  YYLDFV++L     K H LS+VGLSP ET   ILK +  S
Sbjct: 65  ADTRLGSINSMGLPNNGFRYYLDFVSDLAKTGQHKPHFLSVVGLSPTETETILKAINASD 124

Query: 121 YNGLIELNLSCPNVPGKPQIAYDFEMTDLILSEIFSYYQKPLGIKLPFYFDIVHFDQAAT 180
            Y GL+ELNLSCPNVPGKPQIAYDFE TD +L  IF+YY KPLGIKLPFYFDIVHFDQAA
Sbjct: 125 YEGLVRLNLSCPNVPGKPQIAYDFETTDQLLSNIFTYYTKPLGIKLPFYFDIVHFDQAAA 184

Query: 181 IPNKYPLAPINCVNSIGNGLVIDDETVVIKPKNGFGGIGGDPIKPTALANVHAFYKRLNP 240
            IPNKYPL+P+NCVNSIGNGLVI DE V+IKPKNGFGGIGGD+IKPTALANVHAFYKRL P
Sbjct: 185 IPNKYPLSPVNCVNSIGNGLVIKDEQVLIKPKNGFGGIGGDYIKPTALANVHAFYKRLKP 244

Query: 241 SIKIIGTGGVKNGRDAFEHILCGASMVQIGTALQKEGPEIFQRVSRELKEIMADKGYQSL 300
            SI IIGTGGVK GRDAFEHILCGASMVQIGTAL +KGP IF+RV+ELK IM +KGYQSL
Sbjct: 245 SIHIIGTGGVKTGRDAFEHILCGASMVQIGTALHQEGPAIFERVTKELKTIMVEKGYQSL 304

Query: 301 EDFRGQLNY 309
            +DFRG L Y
Sbjct: 305 DDFRGNLRY 313
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1756

A DNA sequence (GBSx1863) was identified in *S.agalactiae* <SEQ ID 5457> which encodes the amino acid sequence <SEQ ID 5458>. This protein is predicted to be beta-lactam resistance factor. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4437(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB89121 GB:AJ277485 beta-lactam resistance factor
            [Streptococcus pneumoniae]
 Identities = 238/410 (58%), Positives = 304/410 (74%)

Query: 1   MALKEIVTAKEFESYSGNYDLQSFMQTPEMAKLLKKRGYDITYMGYQIDGKMEIISIVYTI 60
```

```
              MAL  LT +EF++YS      +SFMQ+ +M  LL+KRG   I Y+   + +G++++ ++VY++
Sbjct: 1      MALTTLTKEEPQTYSDQVSSRSFMQSVQMGDLLEKRGARIVYLALKQEGEIQVAALVYSL 60

Query: 61    PMTGGLHMEVNSGPAHSNSKYLKHFYKELQNYAKSQGALELLIKPYDTYQEFTGEGKPKG 120
              PM GGLHME+NSGP ++     L  FY EL+  YAK  G LELL+KPY+TYQ F  +G P
Sbjct: 61    PMLGGLHMELNSGPIYTQDALPVFYAELKEYAKQNGVLELLVKPYETYQTFDSQGNPID 120

Query: 121   APNTYLIDDLTSIGYHHDGLHIGYPGGEPDWHYVKNLEGITPQNLLKSFSKKGRPLVKKA 180
              A   +I DLT +GY  DGL  GYPGGEPDW Y K+L  +T  ++LLKSFSKKG+PLVKKA
Sbjct: 121   ASKKSIIQDLIDLGYQFDGLTMGYPGGEPDWLYYKDLTRLTEKSLLKSFSKKGKPLVKKA 180

Query: 181   MSFGIKIRVLKREELHIFKDITSSTSDRRDYMDKSLDYYQDFYDSPGDKAEFVIATLNFR 240
              +PGI+++ LKREEL IFK+IT  TS+RR+Y DKSL+YY+  FYD+PG++AEF+IA+LNF
Sbjct: 181   ETFGIRLKKLKREELSIFKNITKETSERRKYSDKSLEYYHHFYDTPGEQAEFLIASLNFS 240

Query: 241   EYDHNLQLNAKKLKEEQITVLDNRHQNNTDSAKYHRQRTELVNQLASLDKRRKEVEPFIQK 300
              +Y   LQ    KLEE +  L     N  S K   Q  E  +Q  + + R+  E    I+K
Sbjct: 241   DYMSKLQSEQSKIEENLDKLRLDLSKNPHSEKKQNQLREYSSQFETFEVRKAEARDLIEK 300

Query: 301   FGNQDVVLAGSLFIYSPKETVLFSGSYTEFNKFYAPAVLQEYVMQEALKRQSTFYNFLG 360
              +G +D+VLAGSLF+Y P+ETV LFSGSYTEFNKFYAPA+LQ+YVM E++KR    YNFLG
Sbjct: 301   YGSEDIVLAGSLFVYMPQETTYLFSGSYTEFNKFYAPALLQKYVMLESIKRGIPKYNFLG 360

Query: 361   IQGNFDGSDGVLRFKQNFNGYIVRKMGTPRYYPNPLKYKSIQLLKKLRR 410
              IQG FDGSDGVLRFKQNFNGYIVRK GTFRY+P+PLKYK+IQLLKKI+ R
Sbjct: 361   IQGIFDGSDGVLRFKQNFNGYIVRKAGTPRYHPSPLKYKAIQLLKKIVGR 410
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5459> which encodes the amino acid sequence <SEQ ID 5460>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2652(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 216/410 (52%), Positives = 291/410 (70%)

Query: 1    MALKELTAKEPESYSGNYDLQSFMQTPEMAKLLKKRGYDITYMGYQIDGKMEIISIVYTI 60
             MAL E++ ++F+ Y +    SF+QT EMA L+ KRG    ++G + DG++++ ++V++
Sbjct: 1    MALIEISQEQFDHYCHSLVHHSPIQTSEMASLMAKRGAKPQFLGLEKDGELKVAAMVFSQ 60

Query: 61   PMTGGLHMSVNSGPAHSNSKYLKHFYKELQNYAKSQGALELLIKPYDTYQEFTGEGKPKG 120
              + GG  ME+N+GP  ++ + L+HFY +L++YAK +  +ELI+KPYD YQ F  +G P
Sbjct: 61   KVAGGWRMELNAGPNTNHPEELEHFYTQLKDYAKQNDVIELILLKPYDNYQSFDTGIPIS 120

Query: 121  APNTYLIDDLTSIGYHHDGLHIGYPGGEPDWHYVKNLEGITPQNLLKSFSKKGRPLVKKA 180
              PNT LI  LT++GY HDGL  GYP GEP DWHYVK LEGI    L +SFSKKG+ L+KKA
Sbjct: 121  RPNTDLISLLTALGYKHDGLKTGYPEGEPVWHYVKKLEGIDSSRLTRSFSKKGKALIKKA 180

Query: 181  MSFGIKIRVLKREELHIFKDITSSTSDRRDYMDKSLDYYQDFYDSPGDKAEFVIATLNFR 240
              +FGIK+R LKR+ELH FK+IT +TSDRRDY+DKSL YYQDFYDSPGD  EF++ATLNF
Sbjct: 181  NTFGIKLRQLKRDELHHFKEITETATSDRRDYLDKSLSYYQDFYDSPGDSCESMVATLNFE 240

Query: 241  EYDHNLQLNAKKLKEEQITVLDNRHQNNTDSAKYHRQRTELVNQLASLDKRRKEVEPFIQK 300
             +Y +NL+    +L   I  +     N  S K   EL +Q  +  R  E   F++ +
Sbjct: 241  DYLNNLKQRQLQLATSINKVKGDLGKNPHSEKKQNRLKELSSQFETPQVRISEALHFLEE 300

Query: 301  FGNQDVVLAGSLFIYSPKETVLFSGSYTEFNKFYAPAVLQEYVMQEALKRQSTFYNFLG 360
             +G +DV LAGSLFIY+ +E VYLFSGSY +FNKFY+PA+LQE+ M +A+ +    YNFLG
Sbjct: 301  YGTKDVFLAGSLFIYTPQEAVYLFSGSYPKFNKFYSPALLQEHAMLKAIHKGIKQYNFLG 360

Query: 361  IQGNFDGSDGVLRFKQNFNGYIVRKMGTPRYYPNPLKYKSIQLLKKILRR 410
```

```
          I G EDGSDGVLRFKQNPNG+I++K GTFR YP P+KY  I+L KK+L R
Sbjct: 361 ITGKFDGSDGVLRFKQNPNGPILQKPGTFRCYPFPIKYHFIRLAKKLLNR 410
```

A related GBS gene <SEQ ID 8895> and protein <SEQ ID 8896> were also identified. Analysis of this protein sequence reveals the following:

```
Homology to resistance proteins
```

The protein has homology with the following sequences in the databases:

```
57.4/74.9% over 409aa                                          Streptococcus
pneumoniae
  GP|7649683| beta-lactam resistance factor Insert characterized

ORF01118(301 - 1530 of 1833)
GP|7649683|emb|CAB89121.1||AJ277485(1  -  410  of  410)  beta-lactam  resistance  factor
{Streptococcus pneumoniae}
%Match = 39.0
%Identity = 57.3  %Similarity = 74.9
Matches = 235  Mismatches = 103  Conservative Sub.s = 72

240       270       300       330       360       390       420       450
IPVNRLLYKASNYVVALRKRNS*LGKDTPMALKELTAKEPESYSGNYDLQSFMQTPEMALKLLKKRGYDITTMGYQIDGK
                                 |||  || :||::||     :||||: :|  ||:|||  | |: : :|:
                                MALTTLTKEEPQTYSDQVSSRSFMQSVQMGDLLEKRGARIVYLALKQEGE
                                          10        20        30        40        50

480       510       540       570       600       630       660       690
MEIISIVVTIPWTGGLHMEVNSGPAHSNSKYLKHFYKELQNYAKSQGALSLLIKPYDTYQEFTGEGKPKGAPNTYLIDDL
::: ::||::|| ||||||:|||| ::    |  || ||: ||| | ||||:|||:||| |  :| |  |    :| ||
IQVAALVYSLPMLGGLHMELNSGPIYTQCDALPVFYAELKEYAKQNGVLELLVKPYETYQTFDSQGNPIDAEKKSIIQDL
        60        70        80        90       100       110       120       130

720       750       780       810       840       870       900       930
TSIGYHHDGLHIGYPGGEPDWHYVKNLEGITPQNLLKSFSKKGRPLVKKAMSFGIKIRVLKREELHIFKDITSSTSDRRD
| :||: |||  |||||||||| | |:|  :| ::|||||||||:|||||| :|||::: |||||| |||:|| ||:||:
TDLGVQFDGLTTGYPGGEPDWLYYKDLTELTEKSLLKSFSKKGKPLVKKAETFGIRLKKLKREELSIFKNTYKETSERRE
        140       150       160       170       180       190       200       210

960       990       1020      1050      1080      1110      1140      1170
YMDKSLDYYQDFYDSFGDKAEFVIATLNPREYDHNLQLNAKKLEEQITVLDNRHQNNTDSAKYHRQRTELVRQLASLDKR
| ||||:||: |||:||::||:||:||| :|   ||    |||| :  |    | | : |  | :|: ::: |
YSDKSLKEYYEHFYDTFGEQAEFLIASLNFSDYMSKLQGEQSKLSENLDKLRLLSKMPSEKKQNQLREYSSQFEYFEVK
        220       230       240       250       260       270       280       290

1200      1230      1260      1290      1320      1350      1380      1410
RKEVEPPIQKPGNQDVVLAGSLFIYSPKETVYLPSGSYTEPNKFYAPAVLQEYVMQEALKRQSTFVXFLGIQGNPDGS G
: |  :|:|:| :|:|||||||:| |:|| ||||||||||||||||:|:||| |::||    | |||||| ||||:|
KAERARDLIEKYGEEDIVLAGSLFVYMPQSTTYLPSGSYTEPNKFYAPALLQKYVMLESIRKGIPKVNFLGIQGIFDSGDG
        300       310       320       330       340       350       360       370

1440      1470      1500      1530      1560      1590      1620      1650
VLKFKQNFNGYIVRKMGTFRYYPNPLKYKSIQLLKKILRRT*KISLHKLIFYAL*KASPISLLLLFIQTIMFVI*RNFIT
|| ||||||||||||| |||||:|:|||||:|||||||: |
VLRFKQNFNGYIVRKAGTFRYHPSPLKYKAIQLLKKIVGR
        380       390       400       410
```

SEQ ID 8896 (GBS198) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 6; MW 48.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 6; MW 73.8kDa).

GBS198-GST was purified as shown in Figure 223, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1757

A DNA sequence (GBSx1864) was identified in *S.agalactiae* <SEQ ID 5461> which encodes the amino acid sequence <SEQ ID 5462>. This protein is predicted to be MurM protein. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4418(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB89539 GB:AJ250767 MurM protein [Streptococcus pneumoniae]
 Identities = 204/410 (49%), Positives = 286/410 (69%), Gaps = 17/410 (4%)

Query: 1   MYRE---ITAVEHDRFVSESNQTNLLQSSNWPKVKDNWGSQLLGFFDGRTQIASASILIK 57
           MYR     I  +E+D+FV E     N+LQSS W KVK +W  + LG ++GE  +A AS+LIK
Sbjct: 1   MYRYQIGIPTLEYDQFVKEHELANVLQSSAWSKVKSDWNHERLGVYEGSNLLAVASVLIK 60

Query: 58  SLPLGFSMLYIPRGPIMDYSNLDIVTKVLKDLKAFGKKQRALFIKCDPLIYLK--MVNAK 115
           SLPLG+ M YIPRGPI+DY + +++  VL+ +K++ + +RA+F+  DP I L    +VN
Sbjct: 61  SLPLGYKMFYIPRGPILDYMDKELLKFVLQSIKSYARSKRAVFVTFDPSICLSQHLVN-- 118

Query: 116 DFENSPDEKEGLIAIDHLQRAGADWIGRTTDLAHTIQPRFQANLYANQFGLDKMSKKTRQ 175
              ++  +  E L  ++ L + G  W+G+TT++  TIQPR QA +Y   F  DK+SK TRQ
Sbjct: 119 --QDKREYPENLAIVEILLGLGVKWSGQTTEMDDTIQPRIQAKIYKENFEEDKLSKSTRQ 176

Query: 176 AIRTSKNKGVDIQFGSHELLEDFAELMKKTEDRKGINLRGIDYYQKLLDTYPNNSYITMA 235
           AIRT++NKG++IQ+G  ELL+ F+ELMKKTE RK I+LR    YY+KLLD +  +SYIT+
Sbjct: 177 AIRTARNKGLSIQYGGLELLDSFSELMKKTEKRKEIHLKNEAYYRKLLDNFKEDSYITLT 236

Query: 236 SLDVAKRLEKIEKECQIAQSERIKS--LELNREKKVKQHQGTIDRLNKEIDFLKEAQKAY 293
           +LDV+KRL ++E+  Q+A+++ ++    E  R  KV+  +  +RL +EIDFL +
Sbjct: 237 NLDVSKRLRELEE--QLAKNKALEEAPTESTRTSKVEAQKKEKERLVEEIDFL-QGVMNM 293

Query: 294 DRDIIPLAATLTLEFGNTSENIYAGMDDYFKSYSAPIYTWFETAQRAFERGNIWQNMGGI 353
           ++  IPLAATL+LEFG TS N+YAGMDD FK Y+API TW+ETA+ AFERG +WQN+GG+
Sbjct: 294 EKSNIPLAATLSLEFGTTSVNLYAGMDDDFKRYNAPILTWYETARYAFERGMVWQNLGGV 353

Query: 354 ENDLSGGLYHFKSKFEPIIEEFIGEFNIPVN---RLLYKASNYVYALRKK 400
           EN L+GGLYHFK KF P IEE++GEF +P +     LL  A ++    LRKK
Sbjct: 354 ENSLNGGLYHFKEKFNPTIEEYLGEFTMPTHPLYPLLRELAIDFRKTLRKK 403
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5463> which encodes the amino acid sequence <SEQ ID 5464>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2239(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 203/399 (50%), Positives = 274/399 (67%), Gaps = 4/399 (1%)
```

```
Query: 5   ITAVEHDRFVSBSNQTNLLQSSNWPKVKDNWGSQLLGFFDGETQIASASILIKSLPLGFS 64
           I+  EHD+FV    Q  LLQSS W KVKDNW  + + F++   Q+A+A+ LI+ LPLGF+
Sbjct: 13  ISPEEHDQFVLAQPQAGLLQSSKWGKVKDNWKHRISFYRNGVQVAAAACLIRKLPLGFT 72

Query: 65  MLYIPRGPIMDYSNLDIVTKVLKDLKAPGKKQRALFIKCDPLIYLKMVNAKDPENSPDEK 124
           M+YIPRGPIMDY+N +++   V+K LK PGK +RALFIK DP + +K      + + S +
Sbjct: 73  MIYIPRGPIMDYANFELLDFVIKTLKTPGKSKRALFIKIDPSLVIKQT--LEGKESKEND 130

Query: 125 EGLIAIDHLQEAGADWTGRTTIDLAHTIQPRPQANLYAMQPGLDKMSKKTRQAIRTSKNKG 184
             L  I  L++ G +W+GRT +L  TIQPR QAN+YA  F  D + KK +Q+IRT+ NKG
Sbjct: 131 VTLSIIAFLKKLGVEWSGRTKELEDTIQPRIQANIYAKDFPDDSLPKKAKQSIRTATNKG 190

Query: 185 VDIQFGSHELLEDPAELMKKTEDRKGINLRGIDYYQKLLDTYPNNSYITMASLDVAKRLE 244
           V++  G  ELL+DP+ LMKKTE+RKGI  LRG  YYQKLL  Y   SYITMASLD+ ++ +
Sbjct: 191 VNVTIGGSELLDDPSALMKKTENRKGILRGKSYYQKLLGIYAQCSYITMASLDLPECKK 250

Query: 245 KIEKECQIAQSERIKSLELNREKVKQHQGTIDRLNKEIDPLKEAQKAYDRDIIPLAATL 304
            + ++   A +E+ +  + ++  KV ++Q TI RL K++  L E Q A  +  IPLAATL
Sbjct: 251 LLIQQLDKALAEQARLTDKSKPSKVAENQKTIARLQKDLTTLSE-QLATQTRIPLAATL 309

Query: 305 TLEPGNTSENIYAGMDDVPKSYSAPIYTWPETAQRAPERGNIWQNMGGIENDLSGGLYHP 364
           TL +G TSEN+YAGMDD +++Y AP+ TW+ETA+ AP+RG  W N+GG+EN   GGLYHP
Sbjct: 310 TLIYGETSENLYAGMDDDYRNYQAPLLTWYETAKEAFKRGCRWHNLGGVENQQDGGLYHP 369

Query: 365 KSKPEPIIEEFTGEFNIPVNRLLYKASNYVYALRKKRNS 403
           K++  P IEEF GEFNIPV  L+   +  Y LRKK  S
Sbjct: 370 KARLNPTIEEFAGEFNIPVG-LVSSLAILTYNLRKKLRS 407
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1758

A DNA sequence (GBSx1865) was identified in *S.agalactiae* <SEQ ID 5465> which encodes the amino acid sequence <SEQ ID 5466>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2669(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1759

A DNA sequence (GBSx1866) was identified in *S.agalactiae* <SEQ ID 5467> which encodes the amino acid sequence <SEQ ID 5468>. This protein is predicted to be beta-lactam resistance factor. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.07   Transmembrane   56 -  72 (  55 -  74)

----- Final Results -----
```

```
bacterial membrane --- Certainty=0.1829(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9625> which encodes amino acid sequence <SEQ ID 9626> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB89120 GB:AJ277484 beta-lactam resistance factor
        [Streptococcus pneumoniae]
 Identities = 166/410 (40%), Positives = 250/410 (60%), Gaps = 10/410 (2%)

Query: 6   MYHVTVGISEKEYDAFAIASSQTNLLHSSKWAQVKSNWQNNRRLGFYKDDQLVAVASILIK 65
           MY   +GI   EYD F        N+L SS W +VKSNWQ+E+ G Y++++L+A ASILI+
Sbjct: 1   MYRYQIGIPTLEYDQFVKEHELANVLQSSAWEHVKSNWQHEKFGVYRSEKLLATASILIR 60

Query: 66  SLPLGFTMLYIPRGPIMDYSNKELVNFVLKTLKNPGRKKRAVFAKPDPALLLRQYHLKEE 125
           +LPLG+ M YIPRGPI+DY +KSL+NF ++++K++ R KRAVF  PDP++ L Q  + +E
Sbjct: 61  TLPLGYKMFYIPRGPILDYGDKSLLNFAIQSIKSYARSKRAVFVTFDPSICLSQSLINQE 120

Query: 126 NVAEEIDESRQAIDNLKSAGAQWIGFTYAISETIQPRPQANIYTKANIEENFPKHTRLRI 185
           E  E+   ID+L+  G +W G T+ + +TIQPR QA IY +   E+   K TK+ I
Sbjct: 121 KT--EFPENLAIIDSLQQMGVRWSGKTERMGDVIQPRIQAKIYKENFEEDKLSKSTKQAI 178

Query: 186 KDAKHRGVQIYRANIDDLPKFATVVALTENRKGVALRNENYFHQLMTIYGEDAYLYLAKV 245
           + A++G++I    ++ L  F+ ++  TE RK + LRNE Y+ +L+  + + AY+ LA +
Sbjct: 179 RTARNKGLEIQYGGLELLDSFSELMKTTEKREIHLRNEAYYKKLLDNFKDKAYITLATL 238

Query: 246 NLPKRLAQFKEQLLQIQKDLSETPGHQKSRLTRLNQQEASVKQYILEFQEFSKKYPD--- 302
           ++ KR  + +EQL +  + L ET + + +R +++  Q+   K+ +LE   F ++Y D
Sbjct: 239 DVSKRSQELEEQLAK-NRALEETFT-ESTRTSKVEAQKKE-KERLLEELTFLQEYIDVGQ 295

Query: 303 -EPVIAGILSIRPGNVLEMLYAGMDDSFRKFYPQYLLNARVFEDAFKNDIVSANLGGVEG 361
            +A  LS+ PG      +YAGMDD F+++     L       AF+  ++  NLGGVE
Sbjct: 296 ARVPLAATLSLEPGTTSVNIYAGMDDDFKRYNAPILTWYETARYAFERGMIWQNLGGVEN 355

Query: 362 SLNDGLTKPKSNFNPMFEEYIGEFNLAINPLLYKLANLAYTIRKKQRHSH 411
           SLN GL  FK  FNP  EEY+GEF +  +P LY L  LA   RK  R  H
Sbjct: 356 SLNGGLYHPKEKFNPTIEEYLGEFTMPTHP-LYPLLRLALDFRKTLRKKH 404
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5469> which encodes the amino acid sequence <SEQ ID 5470>. Analysis of this protein sequence reveals the following:

```
        Possible site: 32

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane   59 -  75 (  59 -  75)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB89120 GB:AJ277484 beta-lactam resistance factor
        [Streptococcus pneumoniae]
 Identities = 166/402 (41%), Positives = 255/402 (63%), Gaps = 5/402 (1%)

Query: 9   KIGISEKEHDSFVKEHQQISVLQSSDWARIKNQWQNERRIGIYKEEKQVASLSLLIKLLPL 68
           +IGI   E+D FVKEH+  +VLQ S W ++K+ WQ+E+ G+Y+EEK +A+ S+LI+ LPL
Sbjct: 5   QIGIPTLKYDQFVKEHELANVLQSSAWEVKSNWQHEKFGVYREEKLLATASILIRTLPL 64

Query: 69  GRSIIYIPRGPVMDYLDRDLVAFTMKTLKDYGKTKKALFIKYDPAILLKQYALGQEEEEX 128
           G  + YIPRGP++DY D++L+ F ++++K Y ++K+A+F+ +DP+I L Q  + QE+ E
```

```
Sbjct: 65  GYKMFYIPRGPILDYGDKSLLNFAIQSIKSYARSKRAVFVTFDPSICLSQSLINQEKTEF 124

Query: 129 PLALAAIKNLQEAGVHWTGLTMEIADSIQPRFQANIYTQENLEMQFPKHTRRLIKDAKQR 188
           P  LA I +LQ+ GV W+G T E+ D+IQFR QA IY +    E +  K T++ I+ A+ +
Sbjct: 125 PENLAIIDSLQQMGVRWSGKTREMGDTIQPRICAKIYKENFEEDKLSKSTKQAIRTARNK 184

Query: 189 GVKTYRVSQSELHKPSKIVSLTEKRKNISLRNEAYPQKLMTTYGDKAYLHLAKVNIPQKL 248
           G++          L  PS+++   TEKRK I LRNEAY++KL+   + DKAY+ LA +++ ++
Sbjct: 185 GLEIQVGTLELLDSPSEILMKKTEKRKEHIQRNEAYYKKLLDNFKDKAYITTALDVSKRS 244

Query: 249 DQYRQQLLLINQDITRTQAHQKKRLKKLSDQKASLERYITE---PEGPTDQYPEEVVVAG 305
            +  +QL   N+ +  T   +  R  K+E QK   ER + E    + + D     V +A
Sbjct: 245 QRLEEQLAK-NRALEETFPT-ESTRTSKVEAQKKEKERLLEELTPLQEYIDVQQARVPLAA 302

Query: 306 ILSISYGNVMEMLYAGMNDDFKKFYPQYLLYPNVFQDAYQDGIIWANMGGVEGSLDDGLT 365
            LS+ +G      +YAGM+DDFK++      L +    + A++ G+IW N+GGVE SL+ GL
Sbjct: 303 TLSLEPGTTSVNIYAGMDDDFKRYNAPILTWVETARYAFERGMIWQNLGGVENSLNGGLY 362

Query: 366 KPKANFAPTIEEFIGEFNLPVSPLYHIANTWYKIRKQLKNKH 407
            PK  F PTIEE++GEF +P  PLY +        RK L+ KH
Sbjct: 363 HPKSKFNPTIEEYLGEFTMPTHPLYPLLRLAIDFRKTLKKGH 404
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 226/407 (55%), Positives = 318/407 (77%), Gaps = 3/407 (0%)

Query: 5   LMYHVTVGISEKEYDAFAIASSQTNLLHSSKWAQVKSNWQNERLGFYKDDQLVAVASILI 64
            L ++  +GISE+E+D+F     Q ++L  S WA++K+ WQNER+G YK+++ VA  S+LI
Sbjct: 4   LTFYAKIGISEEEHDSFVKEHQQISVLQGSDWAKIKNQWQNERIGIYKEEKQVASLSLLI 63

Query: 65  KSLPLGFTMLYIPRGPIMDYSNKELVNFVLKTLKNFGRKKRAVFAKPDPALLLRQYHLKS 124
            K LPLG +++YIPRGP+MDY ++LV F +KTLK++G+  K+A+F K+DPA+LL+QY L +
Sbjct: 64  KLLPLGRSIIYIPRGPVMDYLDRDLVAFTMKTLKDYGKTKKALFIKYDPAILLKQYALSQ 123

Query: 125 ENVAEEIDESRQAIDNLKSAGAQWIGPTYAISETIQPRFQANIYTANIEENFPKHTKPL 184
            E  EE   + AI NL+ AG  W G T  I+++IQPRFQANIYT+ N+E  FPKHT+RL
Sbjct: 124 EE--EEKPIALAAIKNLQEAGVHWTGLTMEIADSIQPRFQANIYTQENLEMQFPKHTRRL 181

Query: 185 IKDAKHRGVQIYRANIDDLPKFATVVALTENRKGVALRNENYPHQLMTIYGEDAYLYLAK 244
            IKDAK RGV+ YR +  +L KF+ +V+LTE RK ++LRNE YF +LMT YG+ AYL+LAK
Sbjct: 182 IKDAKQRGVKTYRVSQSELHKFSKIVSLTEKRKNISLRNEAYFQKLMTTYGDKAYLHLAK 241

Query: 245 VNLPKRLAQFKEQLLQIQKDLSETPSHQKSRLTRLNQQEASVKQYILEPQEPSKKYPDEP 304
            VN+P++L Q+++QL+ I +D++ T +HQK RL +L  Q+ASV++YILEPQEPSKKYPDEP
Sbjct: 242 VNIPQKLDQVRQQLILINQDITRTQAHQKKRLKKLEDQKASLERYITEPEGPTDQYPEEV 301

Query: 305 VIAGILSIRFGNVLEMLYAGMDDSFRKFYPQYLLNARVFEDAFKNDIVSANLGGVEGSLN 364
            V+AGILSI +GNV+EMLYAGM+D F+KFYPQYLL   VF+DA++ I+ AN+GGVEGSL+
Sbjct: 302 VVAGILSISYGNVMEMLYAGMNDDFRKFYPQYLLYPNVFQDAYQDGIIWANMGGVEGSLD 361

Query: 365 DGLTKFKSNFNPMPKEYIGEFNLAINPLLYKLANLAYTIRKKQRHSH 411
            DGLTKFK+NF P  EE+IGEFNL ++P LY +AN  Y IRK+ ++ H
Sbjct: 362 DGLTKFKANFAPTIEEFIGEFNLPVSP-LYHIANTMYKIRKQLKNKH 407
```

SEQ ID 5468 (GBS377) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 65 (lane 4; MW 49kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 71 (lane 4; MW 74kDa).

GBS377-GST was purified as shown in Figure 212, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1760

A DNA sequence (GBSx1867) was identified in *S.agalactiae* <SEQ ID 5471> which encodes the amino acid sequence <SEQ ID 5472>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2073(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9627> which encodes amino acid sequence <SEQ ID 9628> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC76720 GB:AE000446 orf, hypothetical protein [Escherichia coli K12]
 Identities = 127/269 (47%), Positives = 189/269 (70%), Gaps = 1/269 (0%)

Query: 7   SIKLVAVDIDGTLLNSKREITPEVAKAVQEAKSKGVKIVIATGRPIIGVQDLLEELKLNE 66
           +IKL+A+D+DGTLL      I+P V  A+  A+++GV +V+ TGRP  GV + L+EL + +
Sbjct: 2   AIKLIAIDMDGTLLLPDHTISPAVKNAIAAARARGVNVVLTTGRPYAGVHNYLKELHMEQ 61

Query: 67  EGDYVITFNGGLVQDTATGDDIIKETLTYEDYLDFELLARKLGVHMHAITKEGIYTANRD 126
            GDY IT+NG LVQ  A G  + +  L+Y+DY   E L+R++G H HA+ +  +TTANRD
Sbjct: 62  PGDYCITYNSALVQKAADGSTVACTALSYDDYRFLEKLSREVGSHFHALDRTTLYTANRD 121

Query: 127 IGKYTIHEVTLVNMPLFYRTPEEMG-DKEIIKLMMIDQPDILDAAIAKIPKKVLDNYTIV 185
           I  YT+HE  +PL +   E+M  + + +K+MMID+P ILD AIA+IP++V + YT++
Sbjct: 122 ISYYTVHESFVATIPLVFCEAEKMDPNTQFLKVMMIDEPAILDQAIARIPQEVKEKYTVL 181

Query: 186 KSTPFYLEILDKNVNKGTALLHLAEKMGLTVDQTMAIGDEENDRAMLEVVGNPQNGN 245
           KS P++LEIL K VNKGT +  LA+ +G+  ++ MAIGD+END AM+E  G  V M N
Sbjct: 182 KSAPYFLEILDKRVNKGTGVKSLADVLGIKPERIMAIGDQENDIAMIEYAGVGVAMDNAI 241

Query: 246 PELKKIAKYITKSNEESGVAYALREWVIN 274
           P +K++A ++TKSN E GVA+A+ ++V+N
Sbjct: 242 PSVKEVANFVTKSNLEDGVAFAIEKYVLN 270
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3407> which encodes the amino acid sequence <SEQ ID 3408>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3474(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 197/268 (73%), Positives = 235/268 (87%)

Query: 7   SIKLVAVDIDGTLLNSKREITPRVAKAVQRAKSKGVKIVIATGRPIIGVQDLLRELKLNE 66
           SIKLVAVDIDGTLL   R IT +V +AVQEAK++GV +VIATGRPI GV  LLE+L+LN
Sbjct: 2   SIKLVAVDIDGTLLTDDRRITDDVFQAVQEAKAQGVHVVIATGRPIAGVISLLEQLELNH 61

Query: 67  SGDYVITFNGGLVQDTATGDDIIKETLTYRDYLDFELLARKLGVHMHAITKEGIYTANRD 126
           +G++VITFNGGLVQD  TG++I+KE +TY+DYL+ E L+RKLGVHMHAITKEGIYTANR+
Sbjct: 62  KGNHVITFNGGLVQDAETGEEIVKELMTYDDYLETEFLSRKLGVHMHAITKEGIYTANRN 121

Query: 127 IGKYTIHEVTLVNMPLFYRTPEEMGDKEIIKLMMIDQPDILDAAIAKIPKKVLDNYTIVK 186
```

```
              IGKYT+HE TLVNMP+FYRTPEEM +KEIIK+MMID+PD+LDAAI +IP+   D YTIVK
Sbjct: 122 IGKYTVHESTLVNMPIFYRTPEEMTNKEIIKMMMIDSPDLLDAAIKQIPQHFPDKYTIVK 181

Query: 187 STPFYLEILPKNVNKGTALLHLAEKMGLTVDQTMAIGDENDRAMLEVVGNPVVMQNGNP 246
           STPFYLE +PK V+KG A+ HLA+K+GL + QTMAIGD ENDRAMLEVV NPVVM+NG P
Sbjct: 182 STPFYLEFMPKTVSKGNATKHLAKKLGLDMSQTMAIGDAENDRAMLEVVANPVVMENGVP 241

Query: 247 ELKKIAKYITKSNEESGVAYALEKWVIN 274
           ELKKIAKYITKSN +SGVA+A+R+WV+N
Sbjct: 242 ELKKIAKYITKSNNDSGVAHAIRKWVLN 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1761

A DNA sequence (GBSx1868) was identified in *S.agalactiae* <SEQ ID 5473> which encodes the amino acid sequence <SEQ ID 5474>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2360(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07537 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 211/423 (49%), Positives = 285/423 (66%), Gaps = 5/423 (1%)

Query: 3   EKVFRDPVHTYIHVNNQVIYDLINTKEPQRLRRIKQTSTTSFTFHGAEHSRPSHCLGVYS 62
           EKVF+DPVH YIHV +++I+ LI TKEPQRLRR++Q  TT  TFHGAEH+RF+H LGVYS
Sbjct: 12  EKVFKDPVHRYIHVRDELIWALIGTKEPQRLRRVRQLGTTFLTFHGAEHTRFNHSLGVYS 71

Query: 63  LARKVTEIPDEHYSDLWNKNESLLTMAAALLHDIGHGAYSHTFERLFNTDHEAYTQEIIT 122
           + R++ E+F       WN+ E LLT+ AALLHDIGHG +SH+FE++F+TDHE +T+ +I
Sbjct: 72  ITRRIIEVFQGR--PYWNEERRLLTLCAALLHDIGHGPFSHSFEKVFDTDHEEWTRRMIV 129

Query: 123 NPTTEINAILRKVAPDFPDKVASVINHSYPNKQVVQLISSQIDCDRMDYLLRDSYYTAAS 182
            T EI+ +L K+  DFP KVA VI  +YPNK V  +ISSQID DRMDYL RD+YYT  S
Sbjct: 130 GDT-EIFNVLLKMGDDPPQKVADVIEKTYPNKLVTSIISSQIDADRMDYLQRDAYYTGVS 188

Query: 183 YGQPDLTRILRVIRPTDSGIAFARNGMHAVEDYIVSRFQMYMQVYFHPASRAMELLLQNL 242
           YG FD+ RILRV+RP +  +   ++GMHAVEDYI+SR+QMY QVYFHP +R+ E++L  +
Sbjct: 189 YGHFDMERILRVMRPMEDQVVIKQSGMHAVEDYIMSRYQMYWQVYFHPVTRSAEVILSKV 248

Query: 243 LKRARFLFDTHRDFFECTSPNLIPFTDQYDLQDYLALDDGVMNTYFQSWMQADDNILAD 302
            KR + L++      F+Q   +    F    L DYL LD+ +   YFQ W + +D IL+D
Sbjct: 249 FKRVKDLYEQGYK-FKQEPKHFYSLFECNMSLDDLYLRLDESITMYYFQIWQEEDDRILSD 307

Query: 303 LANRFINRKVFKSITPEKSDKKN-LVKMKELVSQVGFDPDYYIGVHANPDLPYDVYRPEH 361
           L  RFINR++FK I F  + + N   ++++L +Q   DP+YY  V ++ DLPYD YRP
Sbjct: 308 LCVRFINRQLFKYISFNENLQMNDWPRLQQLFAQAEIDPEYYLVVDSSSDLPYDFYRPGE 367

Query: 362 SNPRTEIQIIQKNKQLAELSSLSPIVKALTGSNYGDQRFYFPKEMVLTLDSLFSSTKEEPQ 421
            R  I +I  NG+L ELS  S +V+A++G   D + YFP + L+  S     K+E
Sbjct: 368 EEERLPIHLIMPNGKLRELSRESDVVEAISGKKKTDHKLYFPMDCLTDQSDHKEIKQHIL 427

Query: 422 SYI 424
           S +
Sbjct: 428 SLL 430
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5475> which encodes the amino acid sequence <SEQ ID 5476>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2220(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 321/428 (75%), Positives = 379/428 (88%)

Query: 1   MNEKVFRDPVHTYIHVNNQVIYDLINTKEPQRLRRIKQTSTTSFTFHGAEHSRFSHCLGV 60
            MNEKVFRDPVH YIH++N +IYDLINTKEPQRLRRIKQ  TT+FTFHGAEHSRFSHCLGV
Sbjct: 1   MNEKVFRDPVHNYIHIDNPLIYDLINTKEPQRLRRIKQVPTTAFTFHGAEHSRFSHCLGV 60

Query: 61  YELARKVTEIFDEHYSDLWNKNESLLTMAAALLHDIGHGAYSHTFERLFNTDHRAYTQEI 120
            YE+AR+VT IF+E Y+D+WNK+ESL+TM AALLHDIGHGAYSHTFE LF+TDHEA+TQEI
Sbjct: 61  YEIARRVTAIFEEKYADIWNKDESLVTMTAALLHDIGHGAYSHTFEVLFHTDHEAFTQEI 120

Query: 121 ITNPTTEINAILRKVAPDFPDKVASVINESYPNKQVVQLISSQIDCDRMDYLLRDSYYTA 180
            ITNP TEINAIL + APDFPDKVASVINF+YPNKQVVQLISSQIDCDRMDYLLRDSY++A
Sbjct: 121 ITNPETEINAILVRHAPDFPDKVASVINHTYPNKQVVQLISSQIDCDRMDYLLRDSYFSA 180

Query: 181 ASYGQFDLTRILRVIRPTDSGIAFARNGMHAVEDYIVSRFQMYMQVYFHPASRAMELLLQ 240
            A+YGQFDL RILRVIRP + GI F  +GMHAVEDYIVSRFQMYMQVYFHPASRA+EL+LQ
Sbjct: 181 ANYGQFDLMRILRVIRPVEDGIVFEHSGMHAVEDYIVSRFQMYMQVYFHPASRAVELLLQ 240

Query: 241 NLLKRARFLFDTHRDFFEQTSPNLIPFFTDQYDLQDYLALDDGVMFTYFQSWMQADDNIL 300
            NLLKRA+ L+   + +F++T+P LIPFF  + +L DY+ALDDGVMFTYFQ WM ++D+IL
Sbjct: 241 NLLKRAQHLYPEQQAYFQKTAPGLIPFFEKKANLADYIALDDGVMFTYFQVWMASEDHIL 300

Query: 301 ADLANRFINRKVFKSITFEESDKENLVKMKELVSQVGFDPDYYTGVHANFDLPYDVYRPE 360
            +DLA+RFINRK+ KS+TF++  +  L ++++LV  VGFDPDYYTG+H NFDLPYD+YRPE
Sbjct: 301 SDLASRFINRKILKSVTFDQDSQGELERLRQLVESVGFDPDYYTGIHINFDLPYDIYRPE 360

Query: 361 HSNPRTEIQIIQKNGQLAELSSLSPIVKALTGSNYGDQRFVFPKEMLTLDSLFSSTKEEF 420
             NPRT+I+++QK+G LAELS LSPIVKALTG+ YGD+RFVFPKEML LD LF+ +KE F
Sbjct: 361 LENPRTQIEMMQKDSSLAELSQLSPIVKALTGTTYGDRRFVFPKEMLELDDLFAPSKETF 420

Query: 421 QSYITNEH 428
             SYI+N H
Sbjct: 421 MSYISNGH 428
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1762

A DNA sequence (GBSx1869) was identified in *S.agalactiae* <SEQ ID 5477> which encodes the amino acid sequence <SEQ ID 5478>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4789(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5479> which encodes the amino acid sequence <SEQ ID 5480>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3650(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 64/127 (50%), Positives = 89/127 (69%)

Query: 5   MKLEINNNIQIDNETEMIHEIHDCQFIEKGSYVYINYINAKGERVVIKANHEKLIMTRFS 64
           MKL++ N+I+  +ETE+I EIHDC++ EKG Y YL Y N + E+VVIK N  EL M+RFS
Sbjct: 1   MKLQLTNHIRFGDETEIIQEIHDCEWREKGSYQYLIYQNTDKEKVVIKYNETELIMSRFS 60

Query: 65  NPKSVMRPHRETPALVNIPTPLGVQHLITETSHYQFDLSCQRLHINVVLKQTETGDCFAN 124
           NP+S+M+F      L+ +PTP+GVQ  +T+TSHY  D S Q+L ++Y L Q +T   FA+
Sbjct: 61  NPQSIMKFFAGKKVLIALPTPMGVQQFLTDTSHYHLDCSCQKLDLHYHLLQAQTEMLFAS 120

Query: 125 YELRIQW 131
           Y L + W
Sbjct: 121 YHLELSW 127
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1763

A DNA sequence (GBSx1870) was identified in *S.agalactiae* <SEQ ID 5481> which encodes the amino acid sequence <SEQ ID 5482>. This protein is predicted to be cation-transporting ATPase PacL (ctpF). Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.27   Transmembrane  256 - 272 ( 246 - 276)
   INTEGRAL    Likelihood = -9.02   Transmembrane   64 -  80 (  58 -  85)
   INTEGRAL    Likelihood = -8.49   Transmembrane  833 - 849 ( 828 - 855)
   INTEGRAL    Likelihood = -8.17   Transmembrane   89 - 105 (  81 - 107)
   INTEGRAL    Likelihood = -7.48   Transmembrane  864 - 880 ( 860 - 884)
   INTEGRAL    Likelihood = -3.29   Transmembrane  287 - 303 ( 284 - 306)
   INTEGRAL    Likelihood = -2.55   Transmembrane  754 - 770 ( 753 - 773)
   INTEGRAL    Likelihood = -0.85   Transmembrane  695 - 711 ( 694 - 711)
   INTEGRAL    Likelihood = -0.75   Transmembrane  793 - 809 ( 792 - 809)

----- Final Results -----
            bacterial membrane --- Certainty=0.6307(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13439 GB:Z99112 similar to calcium-transporting ATPase
           [Bacillus subtilis]
 Identities = 380/888 (42%), Positives = 545/888 (60%), Gaps = 49/888 (5%)

Query: 10  FYTQGQEEVLTSLESS-REGLSTTEAKNRLEMYGRNELEEGKKRSLIAKFFDQFKDLMII 68
           F+  GQ ++L +  +S ++GL+  E K RL+ +G NKL+KGKK S +  FF QFKD M++
Sbjct: 3   FHEMGQTDLLEATNTSNKQGLTEKEVKKRLDKHGPNKLQEGKKTSALLLFFAQFKDFMVL 62

Query: 69  ILLVAAALSVITEGMHG-LTDALIILAVVILNAAPGVYQESQAEAAIEALKDMSSPIARV 127
```

```
              +LL A  +S    G  G    DA+ I+A+V +N   G +QE +AE +++ALK++S+P
Sbjct: 63  VLLAATLIS----GFLGEYVDAVAIIAIVFVNGILGFFQERRABQSLQALKELSTPHVMA 118

Query: 128 RRDGHTIEVDSKELVPGDLVMLEAGDVVPADLRILEAASLKIEEAALTGESVPVEKDISQ 187
             R+G   ++ SKELVPGD+V   +GD + AD+R++EA SL+IEE+ALTGES+PV K   +
Sbjct: 119 LREGSWTKIPSKELVPGDIVKFTSGDRIGADVRIVEARSLEIEESALTGESIPVVKHADK 178

Query: 188 VVAEDAGIGDRVNMAYQNSNVTYGRGVGVVTNTGMNTEVGKIADMLANADESETPLKQSL 247
             +  D  +GD  NMA+  + VT G G GVV  TGM T +GKIADML +A   TPL++ L
Sbjct: 179 LKKFDVSLGDITNMAPGTIVTRGSGVGVVVGTGMNTAMGKIADMLESAGTLSTPLQRRL 238

Query: 248 VQLSKLLTYLIVIIAVITFLVGIFVRKEGWIEGLMTSVALAVAAIPEGLPAIVTIVLSMG 307
             QL K+L  + +++ V+   VG+ ++        +  V+LAVAAIPEGLPAIVT+ LS+G
Sbjct: 239 EQLGKILIVVALLIVLVVAVGV-IQGHDLYSMFLAGVSLAVAAIPEGLPAIVTVALSLG 297

Query: 308 TKTLAKRNSIVRKLPAVETLGSTEIIASDKTGTLTMNQMTVEKVYT-------------- 353
             + + K+ SIVRKLPAVETLG   II SDKTGT+T N+MTV  V++
Sbjct: 298 VQRMIKQKSIVRKLPAVETLGCASIICSDKTGTMTQNKMTVTHVWSGGKTWRVAGAGYEP 357

Query: 354 NGVLQSSSEEISVDNNTL--------RIMNFSNDTKIDPSGKLIGDPTETALVQPGLDKN 405
             G   + +EISV+ +            + N SN  K D        L GDPTE AL+
Sbjct: 358 KGSFTLNEKEISVNEHKPLQQMLLPGALCNNSNIEKRDGEYVLDGDPTEGALLTAARKGG 417

Query: 406 FDVREVLKNEPRVAELPFDSDRKLMSTTHKESDGRYFIAVKGAPDQLLKRVTKIEDNGLV 465
             F   V  N   + E PPDS RK+M+ I +  D + +I  KGAPD L++R ++I  +G
Sbjct: 418 FSKEFVESNYRVIEEFPFDSARKMMTVIVENQDRKRYIITKGAPDVLMQRSSRIYYDGSA 477

Query: 466 RDITAEDKEAILNTNKELAKQALRVLMMAYK--YETQIPSLETDIVESDLVFSGLVGMID 523
             + E K       + LA QALR + +AY+     + PS+E   E DL   GL G+ID
Sbjct: 478 ALFSNERKAETRAVLRHLASQALRTIAVAYRPIKAGSTPSMRQ--AEKDLTMLGLSGIID 535

Query: 524 PERPEAAEAVRVAKEAGIRPIMITGDHQDTAEAIAKRLGIIDANDTEDHVFTGAELNELS 583
             P RPE  +A++  +EAGI+ +MITGDH +TA+AIAK L ++     +   G  LNELS
Sbjct: 536 PPRPEVRQAIKECREAGIKTVMITGDHVETAKAIAKDLRLLPKS---GKIMDGIMLNELS 592

Query: 584 DEEFQKVFKQYSVVARVSPEHKVRIVKAWQNDGKVVAMTGDGVNDAPSLKTADIGIGMGI 643
             EE   V +   V+ARVSPEHK++IVKA+Q +G +VAMTGDGVNDAP++K ADIG+ MGI
Sbjct: 593 QEELSHVVEDVYVRARVSPEHKLKIVKAYQENGHIVAMTGDGVNDAPAIKQADIGVSMGI 652

Query: 644 TGTEVSKGASDMVLADDNFATIIVAVEEGRKVFSNIQKSIQYLLSANMAEVFTIFFATLL 703
             TGT+V+K AS +VL DDNFATI  A++EGR ++ NI+K I+YLL++N+ E+  + FA LL
Sbjct: 653 TGTDVAKEASSLVLVDDNFATIKSAIKEGRNIYENIRKFIRYLLASNVGEILVMLFAMLL 712

Query: 704 GWDV-LAPVHLLWINLVTDTLPAIALGVEPAEPGVMTHKPRGRQGNFFDGGVMGAIIYQG 762
             + L P+ +LW+NLVTD LPA+ALG++  E  VM  KPR  +  P  +  ++ +G
Sbjct: 713 ALPLPLVPIQILWVNLVTDGLPAMALGMDQPEGDVMKRKPRHPKEGVFARKLGWKVVSRG 772

Query: 763 ILQTILVLGVYGWALMY---PEHAGYRMIHADALTMAFATLGLIQLVHAFNVKSVYQSIF 819
             L I V  +  ++Y   PE+  Y     A T+AFATL L QL+H F+ +S  S+F
Sbjct: 773 FL--IGVATILAFIIVYHRNPENLAY------AGTIAFATLVIAQLIHVFDCRS-ETSVF 823

Query: 820 TVGAFKNRTFNWSIPVAFILLMVTIVVPGFNKLFHVTHLSSTQWLTVV 867
             +  F+N    ++  + +L++V I  P   +FH   ++  W+ V+
Sbjct: 824 SRNPFQNLYLIGAVLSSILLMVVIYVPPLQPIFHTVAITFGDWMLVI 871
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4171> which encodes the amino acid sequence <SEQ ID 4172>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.47   Transmembrane   863 - 879 ( 856 - 883)
   INTEGRAL    Likelihood =-10.08   Transmembrane    64 -  80 (  58 -  86)
   INTEGRAL    Likelihood = -8.97   Transmembrane   256 - 272 ( 249 - 275)
   INTEGRAL    Likelihood = -8.55   Transmembrane    89 - 105 (  81 - 107)
   INTEGRAL    Likelihood = -5.84   Transmembrane   832 - 848 ( 827 - 850)
   INTEGRAL    Likelihood = -3.13   Transmembrane   287 - 303 ( 284 - 307)
   INTEGRAL    Likelihood = -2.66   Transmembrane   762 - 778 ( 761 - 779)
```

```
   INTEGRAL     Likelihood = -0.37   Transmembrane  685 - 701 ( 685 - 701)

  ----- Final Results -----
                  bacterial membrane --- Certainty=0.5989(Affirmative) < succ>
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
    Identities = 735/892 (82%), Positives = 813/892 (90%), Gaps = 1/892 (0%)

Query: 3   KEQKKSLFYTQQQSEVLTSLESSSREGLSTTRAKNRLEMYGRNELEEGKKRSLIAKFFDQF 62
            KEQ+    FYTQ +E VL   LE+SREGL++ +AK RL  YGRNEL+EG+KRSL   KF DQF
Sbjct: 3   KEQRHEAFYTQSEETVLAQLETSREGLTSAQAKERLAEYGRNELDEGEKRSLPMKFLDQF 62

Query: 63  KDLMIILLVAAALSVITEGMHGLTDALIILAVVIVNAAPGVYQEQQAEAAIEALKDMSS 122
            KDLMIILL+VAA LSV+TEGM  GLTDA+IILAVVILVNAAPGVYQEQQAEAAIEALK MSS
Sbjct: 63  KDLMIILLVAALLSVLTEGMEGLTDAIIILAVVILVNAAPGVYQEQQAEAAIEALKSMSS 122

Query: 123 PIARVRRDGHTIEVDSKELVPGDLVMLEAGDVVPADLRLLEAASLKIEEAALTGESVPVE 182
            P+AR+RRDGH  E+DSKELVPGD+V+LEAGDVVPADLRLLEA SLKIEEAALTGESVPVE
Sbjct: 123 PLARIRRDGHVTEIDSKELVPGDIVLLEAGDVVPADLRLLEANSLKIEEAALTGESVPVE 182

Query: 183 KDISQVVAEDAGIGDRVNMAYQNSNVTYGRGYGVVTNTGMYTEVGKIADMLANADESETP 242
            KD+S  V+EDAGIGDRVNM YQNSNVTYGRG GV+TNTGMYTEVG IA MLANADE++TP
Sbjct: 183 KDLSTAVSEDAGIGDRVNMGYQNSNVTYGRGIGVITNTGMYTEVGHIAGMLANADETDTP 242

Query: 243 LKQSLVQLSKLLLTYLIVIIAVITFLVGIPVRKEGWIEGLMTSVALAVAAIPEGLPAIVTI 302
            LKQ+L  LSK+LTY I++IA +TF VG+P+R +  +EGLMTSVALAVAAIPEGLPAIVT+
Sbjct: 243 LKQNLDNLSKLLTYAILVIAAVTFAVGVFLRGQHPLEGLMTSVALAVAAIPEGLPAIVTV 302

Query: 303 VLSMGTKTLAKRNSIVRKLPAVETLGSTEIIASDKTGTLTMNQMTVEKVYTNGVLQSSSE 362
            VLS+GT+ LAKRN+I+RKLPAVETLGSTEIIASDKTGTLTMNQMTVEKVYTNG LQSSS
Sbjct: 303 VLSLGTQVLAKRNAIIRKLPAVETLGSTEIIASDKTGTLTMNQMTVEKVYTNGTLQSSSA 362

Query: 363 EISVDNNFTLRIMNFSNDTKIDPSGKLIGDPTETALVQFGLDKNFDVREVLKNEPRVAELP 422
            +I+ DN TLR+MNF+NDTK+DPSGKLIGDPTETALV+FGLD NFDVRE +  EPRVAELP
Sbjct: 363 DIAPDNTTLRVMNFANDTKVDPSGKLIGDPTETALVEFGLDHNFDVREAMVAEPRVAELP 422

Query: 423 FDSDRKLMSTIHKESDGRYFIAVKGAPDQLLKRVTKIEDNGLVRDITARDKEAILNTMKE 482
            FDSDRKLMSTIHK+DG+YFIAVKGAPDQLLKRVT+IE+NG +R IT  DK+  IL+TNK
Sbjct: 423 FDSDRKLMSTIHKQADGKYFIAVKGAPDQLLKRVTQIEENGQIRPITDADKKTILDTNKS 482

Query: 483 LAKQALRVLMMAYKYETQIPSLETDIVESDLVFSGLVGMIDPERPEAAAVRVAKEAGIR 542
            LAKQALRVLMMAYKY    +P+LET+IVE++LVFSGLVGMIDPERPEAA+AV+VAKEAGIR
Sbjct: 483 LAKQALRVLMMAYKYSDALPTLETEIVEANLVFSGLVGMIDPERPEAAQAVKVAKEAGIR 542

Query: 543 PIMITGDHQDTAEAIAKRLGIIDANDTEDHVFTGAELNELSDEEFQKVFKQYSVVARVSP 602
            PIMITGDHQDTA+AIAKRLGII+  D  DHVFTGAELNELSDEEFQKVFKQYSVVARVSP
Sbjct: 543 PIMITGDHQDTAKAIAKRLGIIE-EDGVDHVFTGAELNELSDEEFQKVFKQYSVVARVSP 601

Query: 603 EHKVRIVKAWQNDGKVVAMTGDGVNDAPSLKTADIGIGMGITGTEVSKGASDMVLADDNF 662
            EHKVRIVKAWQN+GKVVAMTGDGVNDAPSLKTADIGIGMGITGTEVSKGASDMVLADDNF
Sbjct: 602 EHKVRIVKAWQNEGKVVAMTGDGVNDAPSLKTADIGIGMGITGTEVSKGASDMVLADDNF 661

Query: 663 ATIIVAVEEGRKVFSNIQKSIQYLLSANMAEVFTIFFATLLGWDVLAPVHLLWINLVTDT 722
            ATIIVAVEEGRKVFSNIQK+IQYLLSANMAEVFTIF ATL GWDVL PVHLLWINLVTDT
Sbjct: 662 ATIIVAVEEGRKVFSNIQKTIQYLLSANMAEVFTIFLATLFGWDVLQPVHLLWINLVTDT 721

Query: 723 LPAIALGVEPAEPGVMTHKPRGRQSNFFDGGVMGAIIYQGILQTILVLGVYKGWALMYPEH 782
            LPAIALGVEPAEPGVM HKPRGR+S+FFDGGV  AI+YQG  QTILVLGVYG+ALM+PEH
Sbjct: 722 LPAIALGVEPAEPGVMHKPRGRKSSFFDGGVKEAILYQGAFQTILVLGVYGFALMFPEH 781

Query: 783 AGYRMIHADALTMAFATLGLIQLVHAFNVKSVYQSIFTVGAFKNRTFNWSIPVAFILLMV 842
             Y  +HADALTMA+ TLGLIQLVHA+NVKSVYQSIFTVG FKN+ FN+SIPVAF+ LM
Sbjct: 782 TSYHDVHADALTMAYVTLGLIQLVHAYNVKSVYQSIFTVGLFKNKLFNYSIPVAFVALMA 841

Query: 843 TIVVPGFNKLPHVTHLLSSTQWLTVVIGSLLMVVLTEIVKFIQRKLGQDEKAI 894
```

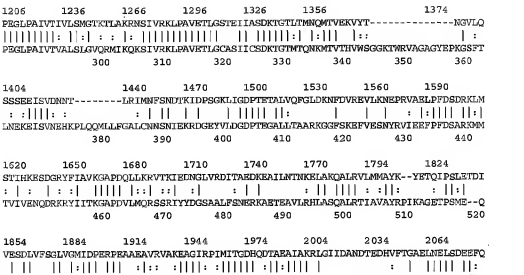
```
          T+VVPGFN+ PHVTHL+ TQWL V+IGSLLMVVL E+VK +QR LGQDEKAI
Sbjct: 842 TVVVPGFNQFPHVTHLTTTQWLVVIIGSLLMVVLVELVKAVQRSLGQDEKAI 893
```

A related GBS gene <SEQ ID 8897> and protein <SEQ ID 8898> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:    -9.88
GvH: Signal Score (-7.5): -6.96
     Possible site: 14
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 9 value: -13.27 threshold:  0.0
   INTEGRAL    Likelihood =-13.27   Transmembrane  256 - 272 ( 246 - 276)
   INTEGRAL    Likelihood = -9.02   Transmembrane   64 -  80 (  58 -  85)
   INTEGRAL    Likelihood = -8.49   Transmembrane  833 - 849 ( 828 - 855)
   INTEGRAL    Likelihood = -8.17   Transmembrane   89 - 105 (  81 - 107)
   INTEGRAL    Likelihood = -7.48   Transmembrane  864 - 880 ( 860 - 884)
   INTEGRAL    Likelihood = -3.29   Transmembrane  287 - 303 ( 284 - 306)
   INTEGRAL    Likelihood = -2.55   Transmembrane  754 - 770 ( 753 - 773)
   INTEGRAL    Likelihood = -0.85   Transmembrane  695 - 711 ( 694 - 711)
   INTEGRAL    Likelihood = -0.75   Transmembrane  793 - 809 ( 792 - 809)
   PERIPHERAL  Likelihood =  1.06     714
modified ALOM score:   3.15

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.6307(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01112(328 - 2901 of 3282)
EGAD|108247|BS1566(3 - 871 of 890) hypothetical protein {Bacillus subtilis} OMNI|NT01BS1841
cation-transporting ATPase PacL GP|2337795|emb|CAA74269.1||Y13937 putative PacL protein
{Bacillus subtilis} GP|2633938|emb|CAB13439.1||Z99112 similar to calcium-transporting
ATPase {Bacillus subtilis} PIR|H69877|H69877 calcium-transporting ATPase homolog yloB -
Bacillus subtilis
%Match = 29.0
%Identity = 43.9  %Similarity = 64.5
Matches = 376  Mismatches = 291  Conservative Sub.s = 176

249        279        309        339        369        396        426        456
GVVLNSETCPHKNRSLFVCGHTKGGKVLLKEQQKKSLFVTQQQEEVLTSLESS-REGLSTIEARNRLEMYGRNELEEGKKR
                                          |:  |: ::| :  :| ::||:  | | ||: :| |||:||||
                                  MKPHEMGQTDLLEATNTSMKQGLTSKEVKKRLDKHGPNELQEGKKT
                                           10         20         30         40

486        516        546        576        606        636        666        696
SLIAKFFDQFKDLMIIILLVAAALSVITEGMEKLTDALIILAVVIINAAAFGVVQESQAEAIEALKDMSSPIARVRRDGH
| :  || ||||:|:::||   | ::|:  :    ||: |:|:| :|  :| :|| :|| :::|||::|:|      |:|
SALLLFFAQFKDFMVLVLL---AATLISGFLGSVVDAVAIIAIVFVNGILGFFQERRAEQSLQALKELSTPHVMALREGS
          60         70         80         90        100        110        120

726        756        786        816        846        876        906        936
TIEVDSKELVPGDLVMLEAGDVVPADLRLLEAASLKIEEAALTGESVPVEKDISQVVAEDAGIGDRVNMAYQNSNVTYGR
  :: ||||||||:| : :|| : ||:|::|| ||:|||:|||||:|| |   ::   |  :||  |||:  : || |
WTKIPSKELVPGDIVKFTSGDRIGADVRIVEARSLEIEESALTGESIPVVKHADKLKKPDVSLGDTTNMAFMGTIVTRGS
          140        150        160        170        180        190        200

966        996        1026       1056       1086       1116       1146       1176
GYGVVTNTGMYTEVGKIADMLANADESETPLKQSLVQLSKLLTVLIVIIAVITFLVGIFVRKEGWIEGLMTSVALAVAAI
| |||  ||| | :|||||| :|    |||:: | || |:|  : ::: |:   ||: ::       ::  |:||||||
GVGVVVGTGMNTAMGKIADMLESAGTLSTPLQRRLEQLGKILIVVALLLTVLVVAVGV- IQGHDLYSMFIAGVSLAVAAI
          220        230        240        250        260        270        280
```



Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1764

A DNA sequence (GBSx1871) was identified in *S.agalactiae* <SEQ ID 5483> which encodes the amino acid sequence <SEQ ID 5484>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
        bacterial cytoplasm --- Certainty=0.2905(Affirmative) < succ>
         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB48940 GB:AJ248283 hypothetical protein [Pyrococcus abyssi]
 Identities = 60/221 (27%), Positives = 100/221 (45%), Gaps = 37/221 (16%)

Query: 33  KIDHLHIA------GDISNHFTKDTLP-FINNLKKH---IKLSYNLGNHDMLDLTE--TE 80
           KID L I       GD+SN+   D +   I+ L      + L    GNHD+  L +
Sbjct: 15  KIDVLKIPDIAIQLGDLSNYGKPDIIRNLISELVTQLDPVPLLVIPGNHDIYGLANDIFAA 74

Query: 81  IQRLDFQTYR------------FDKKMLLAFHGWYDYSFSNN--RDIKDVEKLLKKTWFD 126
           QR +   R                ++  ++    GWYDYS +        KD  ++K  F P
Sbjct: 75  FQRFNKLVKRAGAIPLMEGPLILEEIGIVGVPGWYDYSLAPGYLNWTKDEYEIK-AFGKR 133

Query: 127 RR-----LKRFNNDVTIQASILKRLDEILAKVDSS--NIIIAMHFVPHKQFTMT--HPRF 177
           R      +K   +D  +    L  L++ ++++  S  ++I+A+HF P K        +P
Sbjct: 134 RLEDADYIKSSLSDRELVRWNLNLLEKFISEIRESVNDVILALHFAPFKDSLKYTGNPEI 193

Query: 178 SPFNAFLGSQAYHDLFQKYHIKDVVFGHAHRSFGDVKIGET 218
             F+A++GSQ + +    +++I  +V GH HRS  +  IG+T
Sbjct: 194 DYFSAYMGSQRFGEFALRHNIGLIVHGHTHRSI-EYYIGKT 233
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1765

A DNA sequence (GBSx1872) was identified in *S.agalactiae* <SEQ ID 5485> which encodes the amino acid sequence <SEQ ID 5486>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.18   Transmembrane  173 - 189 ( 173 - 189)

----- Final Results -----
               bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB16056 GB:Z99124 fructose-1,6-bisphosphatase [Bacillus subtilis]
 Identities = 314/642 (48%), Positives = 446/642 (69%), Gaps = 7/642 (1%)

Query: 2   SNFYKLLKEKFPRKEDIVTEMINLEAICQLPKGTEYFISDLHGEYDAVDYLLRTGAGSIR 61
           S +  LL +K+  +E +VTE+INL+AI  LPKGTE+F+SDLHGEY A  ++LR G+G ++
Sbjct: 33  SKYLDLLAQKYDCEEKVVTEIINLKAILNLPKGTEFVSDLHGEYQAFQHVLRNGSGRVK 92

Query: 62  AKLLDCTDWQKIVAVDLDDPCILLYYPKEKLAPDKMNLSASAYKTKLN-EMIPLQIQVLK 120
           K+  D F    I   ++D+   L+YYP+++KL   K+  A     + + E I   I++
Sbjct: 93  EKIRDIPSGV-IYDREIDELAALVYYPEDKLKLIKHDPDAKSALNEWYKETIHRMIKLVS 151

Query: 121 YPSSKYTKSKVRKQLSGKFAYIIEELLAAIDRNPEKKSYPDTIIEKLFELDQVSDLLIVL 180
           Y SSKYT+SK+RK L  +FAYI EELL + ++   K+ Y+  II+++ EL Q + LI  L
Sbjct: 152 YCSSKYTRSKLRKALPAQFAYITEELLYKTEQAGNKEQYYSEIIDQIIELGQADKLITGL 211

Query: 181 SQTIQVLIIDHLHVVGDIYDRGRYPDRILNRLMAFPNLDIQWGNHDVTWMGAASGSYLCM 240
           + ++Q L++DHLHVVGDIYDRG  PDRI+  L+ + ++DIQWGNHDV W+GA SGS +C+
Sbjct: 212 AYSVQRLVVDHLHVVGDIYDRGPQPDRIMEELINYHSVDIQWGNHDVLWIGAYSGSKVCL 271

Query: 241 VNVIRIRARYNNITLIKDRVGINLRRLVDYSRRYYEPLPSFVPILDCEEMTHPDELDLLN 300
           N+IRI ARY+N+ +IED YGINLR L++ + +YY+  P+F P  D  E    DE+   +
Sbjct: 272 ANIIRICARYDMLDIIEDVYGINLRPLLNLAEKYYDDNPAFRPKAD--ENRPEDEIKQIT 329

Query: 301 MIQQATAILQFKLEAQLIDRRPEPQMHNRQLINQVNYKDLSISIKEVVNQLKDFNSRCID 360
            I QA A++QFKLE+ +I RRP F M  R I+ +++Y    I++    +QL++     I+
Sbjct: 330 KIHQATAMIQFKLESPIIKRRPNFNMEERLLLEKIDYDKNEITLNGKTVQLENTCFATIN 389
```

```
Query: 361 SKNPSRLTSEKEELLQQLMIAFQTSRSLKKHIDFLPEKGSMYLTYNDNLLFHCCIPMESN 420
           + P +L  EE E++ +L+ + Q SE L +H++F+ +KGS+YL YN NLL HGCIP+  N
Sbjct: 390 PEQPDQLLEKEAEVIDKLLFSVQHSEKLGKHMNFMMKKGSLYLKYNGNLLIHGCIPVDEN 449

Query: 421 GDFKSFKIAGKTYQGRDLLDLFESQIRLAYARPEKHDDLATDIIWYLWCGSNSSLPGKNA 480
           G+ ++  I  K Y GR+LLD+PE  +R A+A PE+ DDLATD+ WYLW GE SSLPGK A
Sbjct: 450 GNMETMMIEDKPYAGRELLDVFERFLRSAFAHPERTDDLATDMAWYLWTGEYSSLPGKRA 509

Query: 481 MTTFERYYVSDKVTMQRFKNPYFKLRDKDDICTALLQRFDL-PKPGHIVNGHTPVKKKNG 539
           MTTFERY++ +K TH+E+KNPY+ LR+ +  C  +L EF L P  GHI+NGHTPVKE  G
Sbjct: 510 MTTFERYFIKSKETHKSKKNPYYYLRENGATCRNILAEFGLNPDEGHIINGHTPVKEIEG 569

Query: 540 EQPIKANGKMVIDGGFAKGYQKNTGLAGYTLIYNSYGIQLISHLPFTSIEEVLSGTNYI 599
           E PIKANGKM+VIDGGF+K YQ  TG+AGYTL+YNSYG+QL++H  F S  EVLS    +
Sbjct: 570 EDPIKANGKMIVIDGGFSKAYQSTTGIAGYTLLYNSYGMQLVAHKHFNSKAEVLSYGTDV 629

Query: 600 IDTKRLVEEAKDRILVKDTTIGQKLTKEIKDLDHL--YRHFQ 639
           +  KRLV++  +R  VK+T +G++L +E+  L+ L  YR+ +
Sbjct: 630 LTVKRLVDKELERKKVKETNVGEELLQEVAILESLREYRYMK 671
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5486 (GBS197) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 168 (lane 17 & 18; MW 89kDa) and in Figure 169 (lane 2; MW 89kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 6; MW 99kDa).

Purified Thio-GBS197-His is shown in Figure 244, lane 6.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1766

A DNA sequence (GBSx1873) was identified in *S.agalactiae* <SEQ ID 5487> which encodes the amino acid sequence <SEQ ID 5488>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2433(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12719 GB:Z99108 alternate gene name: ypaP-similar to
        hypothetical proteins [Bacillus subtilis]
 Identities = 176/367 (47%), Positives = 240/367 (64%), Gaps = 6/367 (1%)

Query: 3   IKAEIQKLAKEIGISKIGFTTADNFDYLEKSLRASVEEGRNSGFEHKVIRDRIYPERLLE 62
           +K E+ + AK IG+ KIGFTTAD FD L+  L      G  SGFE   IE R+ P+ LL
Sbjct: 55  LKEELIEYAKSIGVDKIGFTTADFDSLKDRLILIQESLGYLSGFEEPDIEKRVTPKLLLP 114

Query: 63  SAKTIISIGVAYPHKLPQQPQKT-SYKRGKITPNSWGLDYHYVVGEKLDRLSKGIEELCR 121
            AK+I++I +AYP ++   P+ T + +RG      SWG DYH V+ EKLD L   ++
Sbjct: 115 KAKSIVAIALAYPSRMKDAPRSTRTERRGIFCRASWGKDYHDVLREKLDLLEDFLKSKHE 174

Query: 122 DFPLQQKAMVDTQALVDTAVAQRAGIGFIGKNGLVISKEYGSYMPLGSLITNLEIEPDKP 181
           D  ++ K+MVDTQ L D AVA+RAGIGF  KN ++ + EYGSY++L E+ITN+. EPD P
Sbjct: 175 D--IRTKSMVDTGELSDRAVARRAGIGPSAKNCMITTPEYGSYVYLASMITNIPPEPDVP 232
```

```
Query: 182 VDYDCGDCRRCLDACPTSCLIGDGSMNAKRCLSFQTQDKGMMDIEFRKKIKTVIYGCDIC 241
           ++  CG C +CLDACPT  L+  G +NA+RC+SF TQ KG +  EFR KI   +YGCD C
Sbjct: 233 IRDMCGSCTKCLDACPTGALVNPGQLNAQRCISFLTQTRGFLPDEFRTKIGNRLYGCDTC 292

Query: 242 QICCPYNKGINNPLATEI--DPELAQPELIPFLSLSNGQPKEKPGMIAGSWRGKNILQRN 299
           Q  CP NKG +  L  E+  DPE+A+P  L P L++SN +PKEKPG ++GSWRGK  +QRN
Sbjct: 293 QTVCPLNKGKDPHLHFEMEMDPSIAKPLLKPLLAISNRERPKEKPGHVSGSWRGKKPIQRN 352

Query: 300 AIYALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEILEPMSNLTLKDE 359
           AI+ALA+  D +A+ +L E++ K+  P+    TA WA+G+I       E LE       KDE
Sbjct: 353 AILALAHFKDASALPELTELMHKDPRPVIRGTAAWAIGKIGDPAYAEELEKALEKE-KDE 411

Query: 360 DSRKELE 366
           +++ E+E
Sbjct: 412 EAKLEIE 418
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5489> which encodes the amino acid sequence <SEQ ID 5490>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3337(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 363/374 (97%), Positives = 367/374 (98%)

Query: 1   MDIKAEIQKLAKEIGISKIGFTTADNFDYLEKSLRASVEBGRNSGFEHKVIEDRIYPERL 60
           M IKAEI+ LAKEIGISKIGFTTADNFDYLEKSLRASVEBGRNSGFEHKVIEDRIY ERL
Sbjct: 18  MTIKAEIKALAKEIGISKIGFTTADNFDYLEKSLRASVEBGRNSGFEHKVIEDRIYTERL 77

Query: 61  LESAKTIISIGVAYPHKLPQQPQKTSYKRGKITFKSWGLDYHYVVGEKLDRLSKGIEELC 120
           LESAKTIISIGVAYPHKLPQQPQKT YKRGKITF+SWGLDYHYVVGEKLDRLSKGIEELC
Sbjct: 78  LESAKTIISIGVAYPHKLPQQPQKTPYKRGKITPSSWGLDYHYVVGEKLDRLSKGIEELC 137

Query: 121 RDFPLQQKAMVDTGALVDTAVAQRAGIGFIGKNGLVISKEYGSYMPLGELITNLEIEPDK 180
           RDFPLQQKAMVDTGALVDTAVAQRAGIGFIGKNGLVISKEYGSYMPLGELITNLEIEPDK
Sbjct: 138 RDFPLQQKAMVDTGALVDTAVAQRAGIGFIGKNGLVISKEYGSYMPLGELITNLEIEPDK 197

Query: 181 PVDYDCGDCRRCLDACPTSCLIGDGSMNAKRCLSFQTQDKGMMDIEFRKKIKTVIYGCDI 240
           PVDYDCGDCRRCLDACPTSCLIGDGSMNAKRCLSFQTQDKGMMDIEFRKKIKTVIYGCDI
Sbjct: 198 PVDYDCGDCRRCLDACPTSCLIGDGSMNAKRCLSFQTQDKGMMDIEFRKKIKTVIYGCDI 257

Query: 241 QQICCPYNKGINNPLATEIDPELAQPELIPFLSLSNGQPKEKPGMIAGSWRGKNILQRNA 300
           QQICCPYNKGINN  ATEIDPELAQPELIPFLSLSNG+PKEKPGMIAGSWRGKNILQRNA
Sbjct: 258 QQICCPYNKGINNSPATEIDPELAQPELIPFLSLSNGKPKEKPGMIAGSWRGKNILQRNA 317

Query: 301 IIALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEILEPMSNLTLKDED 360
           IIALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEILAPMSHLTLKDED
Sbjct: 318 IIALANAHDKTAVVKLIEIIDKNNNPIHTATAIWALGEIVKKPNDEILAPMSHLTLKDED 377

Query: 361 SRKELELIRHKWQF 374
           SRKELELIRHKWQF
Sbjct: 378 SRKELELIRHKWQF 391
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1767

A DNA sequence (GBSx1874) was identified in *S.agalactiae* <SEQ ID 5491> which encodes the amino acid sequence <SEQ ID 5492>. This protein is predicted to be peptide chain release factor 2 , fragment (prfB). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4903(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC67303 GB:AF017113 putative peptide chain release factor RF-2
            [Bacillus subtilis]
 Identities = 194/336 (57%), Positives = 251/336 (73%), Gaps = 2/336 (0%)

Query: 2   EEEIALLENQMTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE- 60
            E  IA L+ QM +P+FWND   AQ    E N LK   +++  + E  +E ++  ++L E
Sbjct: 30  EARIAELDEQMADPEFWNDQQKAQTVINEANGLKDYVNSYKKLNESHEELQMTHDLLKEE 89

Query: 61  -DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILEIHPGSGGTEAQDWGDLLLR 119
             D  L+ ELE+ L  L K    +E+ LLLSEPYD NNAILE+HPG+GGTE+QDWG +LLR
Sbjct: 90  PDTDLQELEKELKSLTKEPNEFELQLLLSEPYDKNNAILELHPGAGGTESQDWGSMLLR 149

Query: 120 MYTRFGNANGFKVEVLDYQAGDEAGIKSVTLSFEGPNAYGLLKSEMGVHRLVRISPFDSA 179
            MYTR+G   GFKVE LDY  GDEAGIKSVTL  +G NAYG LK+E GVHRLVRISPFDS+
Sbjct: 150 MYTRWGERRGFKVETLDYLPGDEAGIKSVTLLIKGHNAYGYLKAEKGVHRLVRISPFDSS 209

Query: 180 KRRHTSFASVEVMPELDDTIEVEVRDDDIKMDTFRSGGAGGQNVNKVSTGVRLTHIPTGI 239
             RRHTSF S EVMPE +D I+++R +DIK+DT+R+ GAGGQ+VN   + VR+TH+PT +
Sbjct: 210 GRRHTSFVSCEVMPEFNDEIDIRTEDIKVDTYRASGAGGQHVNTTDSAVRITHLPTNV 269

Query: 240 VVSSTVDRTQYGNRDRAMKMLQAKLYQLEQEKKAQEVDALKGDKKSITWGSQIRSYVFTP 299
            VV+   +R+Q  NR+RAMKMLQAKLYQ   E++  E+D +KG+KKSI WGSQIRSYVF P
Sbjct: 270 VVTCQTERSQIKNRERAMKMLKAKLYQRRIEEQQAELDEIKGEQKEIGWGSQIRSYVFHP 329

Query: 300 YTMVKDHRTNFELAQVDKVMDGEINGFIDAYLKWRI 335
            Y+MVKDHRTN E+  V  VMDG+I+ FIDAYL+ ++
Sbjct: 330 YSMVKDHRTNVEMGNVQAVMDGDIDTFIDAYLRSKL 365
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5493> which encodes the amino acid sequence <SEQ ID 5494>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4779(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 334/337 (99%), Positives = 336/337 (99%)

Query: 1   MEEEIALLENQMTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE 60
           +EEEIALLEN MTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE
Sbjct: 1   LEEEIALLENHMTEPDFWNDNIAAQKTSQELNELKGKYDTFHNMQELSDETELLLEMLDE 60

Query: 61  DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILEIHPGSGGTEAQDWGDLLLRM 120
           DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILEIHPGSGGTEAQDWGDLLLRM
Sbjct: 61  DDSLKEELEENLMQLDKIMGAYEMTLLLSEPYDHNNAILEIHPGSGGTEAQDWGDLLLRM 120
```

```
Query: 121 YTRPGNANGFKVRVLDYQAGDEAGIKSVTLSFEGPNAYGLLKSEMGVHRLVRISPFDSAK 180
           YTRPGNANGFK+EVLDYQAGDEAGIKSVTLSFEGPNAYGLLKSEMGVHRLVRISPFDSAK
Sbjct: 121 YTRPGNANGFKIEVLDYQAGDEAGIKSVTLSFEGPNAYGLLKSEMGVHRLVRISPFDSAK

Query: 181 RRHTSFASVEVMPELDDTIEVEVRDDDIKMDTFRSGGAGSQNVNKVSTGVRLTHIPTGIV 240
           RRHTSFASVEVMPELDDTIEVEVRDDDIKMDTFRSGGAGSQNVNKVSTGVRLTHIPTGIV
Sbjct: 181 RRHTSFASVEVMPELDDTIEVEVRDDDIKMDTFRSGGAGSQNVNKVSTGVRLTHIPTGIV 240

Query: 241 VSSTVDRTQYGNRDRAMKMLQAKLYQLEQKKAQEVDALKGDKKEITWGSQIRSYVFTPY 300
           VSSTVDRTQYGNRDRAMKMLQAKLYQLEQKKAQEVDALKGDKKEITWGSQIRSYVFTPY
Sbjct: 241 VSSTVDRTQYGNRDRAMKMLQAKLYQLEQKKAQEVDALKGDKKEITWGSQIRSYVFTPY 300

Query: 301 TMVKDHRTNFELAQVDKVMDGEINGFIDAYLKWRIED 337
           TMVKDHRTNFELAQVDKVMDGEINGFIDAYLKWRIED
Sbjct: 301 TMVKDHRTNFELAQVDKVMDGEINGFIDAYLKWRIED 337
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1768

A DNA sequence (GBSx1875) was identified in *S.agalactiae* <SEQ ID 5495> which encodes the amino acid sequence <SEQ ID 5496>. This protein is predicted to be cell-division ATP-binding protein (ftsE). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3928(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC67262 GB:AF017113 cell division ATP-binding protein [Bacillus subtilis]
 Identities = 138/228 (60%), Positives = 179/228 (77%)

Query: 3   LIEMSGVTKKYRRSTTALRNLNLSIQQGEFVVLVGPSGAGKSSLIRLLYREEKLSSGRLK 62
           +IEM  V K Y    AL  ++++I  GEFVV+VGPSGAGKS+ I+++YREEK + G++
Sbjct: 1   MIEMKEVYKAYPNGVKALANGISVTIHPGEFVVVGPSGAGKSTFIRMIYREEKPTKGQIL 60

Query: 63  VGSEPNLNKLKRRQIPILRRSIGVVFQDYKLLPTKTVYENVAFAMQVIGAKRRHIKKRVPE 122
           +   +L  +K ++IP +RR IGVVFQD+KLLP  TV+ENVAFA++VIG +   IKKRV E
Sbjct: 61  INHKDLATIKKEKEIPFVRRKIGVVFQDFKLLPKLTVFENVAFALEVIGEQPSVIKKRVLE 120

Query: 123 VLELVGLKHKMRSPPTQLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPEIAWEIMHLL 182
           VL+LV LKHK R FP QLSGGEQQRV+IAR+IVNNP ++IADEPTGNLDP+ +WE+M  L
Sbjct: 121 VLDLVQLKHKARQFPDQLSGGEQQRVSIARSIVNNPDVVIADEPTGNLDPDTSWEVMKTL 180

Query: 183 ERINLQGTTVLMATHNSQIVNTLRHRVIKIEAGSVIRDERKGEYGYHD 230
           E IN +GTTV+MATHN +IVNT++ RVI IE G ++RDE +GEYG +D
Sbjct: 181 EEINNRGTTVVMATHNKEIVNTMKKRVIAIEDGIIVRDESRGKYGSYD 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5497> which encodes the amino acid sequence <SEQ ID 5498>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3728(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 191/230 (83%), Positives = 214/230 (93%)

Query: 1   MALIEMSGVTKKYRRSTTALRNLNLSIQQGEFVYLVGPSGAGKSSLIRLLYRREKLSSGR 60
           MALIEMSGVTKKYRRSTTALR++N+S+ QGEFVYLVGPSGAGKS+ I+LLYREE+L+SG+
Sbjct: 1   MALIEMSGVTKKYRRSTTALRDVNVSVNQGEFVYLVGPSGAGKSTFIKLLYREEQLTTGK 60

Query: 61  LKVGEFNLNKLKRRQIPILRRSIGVVFQDYKLLPTKTVYENVAFAMQVIGAKRRHIKKRV 120
           L VGEFNL KLK R +PILRR IGVVFQDYKLLP KTV+ENVA+AM+VIG KRRHIKKRV
Sbjct: 61  LYVGEFNLTKLKARDVPILRRHIGVVFQDYKLLPRKTVFENVAYAMEVIGEKRRHIKKRV 120

Query: 121 PEVLELVGLKHKMRSFPTQLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPEIAWEIMH 180
           PEVL+LVGLKHKMRSFP+QLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPEI+WSIM
Sbjct: 121 PEVLDLVGLKHKMRSFPSQLSGGEQQRVAIARAIVNNPKLLIADEPTGNLDPEISWEIMQ 180

Query: 181 LLERINLQGTTVLMATHNSQIVNTLRHRVIEIEAGSVIRDEEKGEYGYHD 230
           LLERIN+QGTT+LMATHNS IVNT RHRV+ IE G ++RDEEKG+YGY D
Sbjct: 181 LLERINVQGTTILMATHNSHIVNTFRHRVVAIEDGRIVRDEEKGDYGYDD 230
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1769

A DNA sequence (GBSx1876) was identified in *S.agalactiae* <SEQ ID 5499> which encodes the amino acid sequence <SEQ ID 5500>. This protein is predicted to be ftsE protein (ftsX). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.77   Transmembrane  296 - 312 ( 291 - 322)
   INTEGRAL    Likelihood = -9.24   Transmembrane  203 - 219 ( 198 - 228)
   INTEGRAL    Likelihood = -6.16   Transmembrane   49 -  65 (  40 -  68)
   INTEGRAL    Likelihood = -3.40   Transmembrane  255 - 271 ( 252 - 273)

----- Final Results -----
               bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9629> which encodes amino acid sequence <SEQ ID 9630> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC67264 GB:AF017113 cell division protein [Bacillus subtilis]
 Identities = 112/311 (36%), Positives = 182/311 (58%), Gaps = 31/311 (9%)

Query: 27  RHFWESLKNLKRNPWMTFASVTSVTITLLLVGLFSSVLLNVKKLVTDVSGNFTISAFLNV 86
           RH  ES K+L RN WMTFAS++VT+TL+LVG+F  ++LN+  + T+          I  +++
Sbjct: 7   RHLRESPKSLGRNTWMTFASISAVTVTLILVGVFLVIMLNLANNMATNAEKQVEIKVLIDL 66

Query: 87  DGTDAQKQVKDKDGKLKDNPDYHKVYDKIRISGVEKVTYSGKAEQLKEVQKEYGSDVID 146
            +           D K +D      K+ + IK + G++ VT+SGK ++L ++    +G
Sbjct: 67  TA----------DQKAQD-----KLQNDIKELKGIQSVTFSSKKEKELDQLVDSFGDSGKS 111

Query: 147 DTYKDA---LLDVYVVGTGSAKVSKSVSEAIGRIEGV---DVTKEPIDST-KLGSNLTDNI 199
            T KD    L D +VV T+    + +V++ I +++ V    Y KE +    K+  ++ NI
Sbjct: 112 LTMKDQENPLNDAFVVKTTDPHDTPNVAKKIEKMDHVYKVTYGKEEVSRLFKVVGVSRNI 171

Query: 200 RIWGFGGVALLIVL---AIFLISNTIRMSIMSRRTDIEIMRLVGAKNSYIRGPFFPEGAW 256
```

```
                G+AL+I L   A+FLISNTI+++I +RR +IEIM+LVGA N +IR PFF EG
Sbjct: 172 ------GIALIIGLVPTAMFLISNTIKITIPARRKRIRIMKLVGATNWFIRWPFFLEGLL 225

Query: 257 VGILGAIVPSLIFYFGYQFVFNKFNPKFETSHVSLYFMDIMVPAIIGGMVIIGIIIGSLG 316
           +G+ G++P +    YQ+V    PK + S VSL P +  V +  ++ IG +IG  G
Sbjct: 226 LGVFGSVIPIALVLSTYQYVIGWVVPKVQGSPVSLLFYNPFVPQVSLVLIAIGAVIGVWG 285

Query: 317 SVLSMRRYLKI 327
           S+ S+R++L++
Sbjct: 286 SLTSIRKFLRV 296
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5501> which encodes the amino acid sequence <SEQ ID 5502>. Analysis of this protein sequence reveals the following:

```
      Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.70   Transmembrane  195 - 211 ( 189 - 219)
   INTEGRAL    Likelihood = -6.74   Transmembrane   39 -  55 (  30 -  58)
   INTEGRAL    Likelihood = -5.52   Transmembrane  294 - 310 ( 288 - 314)
   INTEGRAL    Likelihood = -1.49   Transmembrane  246 - 262 ( 245 - 263)

----- Final Results -----
              bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC67264 GB:AF017113 cell division protein [Bacillus subtilis]
 Identities = 117/311 (37%), Positives = 184/311 (58%), Gaps = 19/311 (6%)

Query: 11  MIRYFFRHIWESIKNLKRNFWMTFASVSMVAVTLLTVGVFAATLLNIQRVASGVENNVHI 70
           MI+   RH+ SS K+L RN WMTFAS+S V VTL LVGVF   +LN+  +A+  E  V I
Sbjct: 1   MIKILGRHLRSSFKSLGRNTWMTFASISAVTVTLILVGVFLVIMLNLNNMATNAEKQVEI 60

Query: 71  NTYLQVDSTPAAKVIQNTAGEPVNNDNYHSVYDKIAQIKGVKKITPSSKDEQLKKLQETL 130
            + + +   A+       + + ND       I ++KG++ +TPSSK++L +L ++
Sbjct: 61  KVLIDLTADQKAQ-------DKLCND--------IKELKGIQSVTPSSKEKELDQLVDSF 105

Query: 131 GDVWN---MYDQETNPLQDIYLIETQTPKQVKAITKKIRTIEGVEAADYGGINSDKLFKF 187
           GD      M DQ+ NPL D ++++T  P     + KKI  ++ V    YG     +LFK
Sbjct: 106 GDSGKSLTMKDQE-NPLNDAFVVKTTDPHDTPNVAKKIEKMDHVYKVTYGKESVSRLFKV 164

Query: 188 STLIQTWGLIGTAMLLFVAVFLISNTIRMTIMSRKRDIEIMRLVGAKNSYIRGPFFFEGA 247
            + +  G+      L+F A+FLISNTI++TI +R++++IEIM+LVGA N +IR PFF EG
Sbjct: 165 VGVSRNIGIALIIGLVFTAMFLISNTIKITIFARRKEIEIMKLVGATNWFIRWPFFLEGL 224

Query: 248 MVGLLGAVLPSLLIYYGYDLVYKHPAQELQRNNLSMYPLDPVVYYLIGALFVIGIMIGSL 307
            +G+ G+V+P  L+    Y V      ++Q + +S+ P +P+V+ +  L  IG +IG
Sbjct: 225 LLGVFGSVIPIALVLSTYQYVIGWVVPKVQGSPVSLLPYNPFVPQVSLVLIAIGAVIGVW 284

Query: 308 GSVLSMRRYLK 318
           GS+ S+R++LK
Sbjct: 285 GSLTSIRKFLR 295
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 173/318 (54%), Positives = 238/318 (74%), Gaps = 5/318 (1%)

Query: 13  MKRRSNMVIMIN-FFRHFWESLKNLKRNFWMTFASVTSVTITLLLVGLFSSVLLNVEKLT 71
           MK++E MV MI  FFRH WES+KNLKRNFWMTFASV+ V +TL LVG+F++ LLN++++
Sbjct: 2   MKKKEIMVTMIRYFFRHIWESIKNLKRNFWMTFASVSMVAVTLTLVGVFAATLLNIQRVA 61

Query: 72  TDVSGNFTISAFLNVDSTDAQKQVKDKDGKLKDNPDYHKVYDKIKRISGVEKVTYSSKAE 131
           + V  N  I+ +L VDSTDA K +++  G+  +N +YH VYDKI +I GV+K+T+SSK E
Sbjct: 62  SGVENNVHINTVLQVDSTDAAKVIQNTAGEPVNNDNYHSVYDKIAQIKGVKKITPSSKDE 121
```

```
Query: 132 QLKEVQKSYGSDVID--DTYKDALLDVYVVGTSSAKVSKSVSEAIGRIBGVDVTKEP-ID 188
           QLK++Q+  G DV +  D   + L D+Y++ T + K  K++++ I  IBGV+      I+
Sbjct: 122 QLKKLQETLG-DVWNMYDQDTNPLQDIYLIETQTPKQVKATTKKIRTIBGVEAADYGGIN 180

Query: 189 STKLSNLTDNIRIWGPGGVALLIVLAIFLISNTIRMSIMSRRTDIEIMRLVGAKNSYIRG 248
           S KL   +  I+ WG  G A+L+ +A+FLISNTIRM+IMSR+  DIEIMRLVGAKNSYIRG
Sbjct: 181 SDKLFKFSTLIQTWGLIGTAMLLFVAVFLISNTIRMTIMSKRKDIEIMRLVGAKNSYIRG 240

Query: 249 PFFFBGAWVGILGAIVPSLIFYFGYVQFVFNKFNPKPETSHVSLYFMDIMVPAIIGGMVII 308
           PFFFBGAWVG+LGA++PSL+ Y+GY  V+  F  + + +++S+YP+D  V  +IG + +I
Sbjct: 241 PFFFBGAWVGLLGAVLPSLLIYYGYDLVYKHFAQELQRNNLSNYPLDPYVYYLIGALFVI 300

Query: 309 GIIIGSLGSVLSMRRYLK 326
           GI+IGSLGSVLSMRRYLK
Sbjct: 301 GIMIGSLGSVLSMRRYLK 318
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1770

A DNA sequence (GBSx1877) was identified in *S.agalactiae* <SEQ ID 5503> which encodes the amino acid sequence <SEQ ID 5504>. This protein is predicted to be carboxymethylenebutenolidase-related protein. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF10898 GB:AE001979 carboxymethylenebutenolidase-related
          protein [Deinococcus radiodurans]
 Identities = 65/183 (35%), Positives = 98/183 (53%), Gaps = 3/183 (1%)

Query: 56  SKGKVKANIIFYQGALVEEEAYSQLARDLADKGDNTYILKTPLNLPVLSPHKATIINQN 115
           +  +VK  ++FY G  V  +AY  L R LA +G  T I   PL+L +     +A+ +I +
Sbjct: 100 ASAEVKTLLVFYPGGRVRPQAYEWLGRALAVRGVQTVIPAFELDLAITGTERAEGLIARY 159

Query: 116 HL-TNVYLAGHSLGGVVASQNAKVAP--VRGLLILLASYPSRKSDLSHKNLRVLSITASND 172
               V LAGHSLGG VA+Q A + P  + GL+LLA+YP+   +L      LS+ A  D
Sbjct: 160 GAGKRVVLAGHSLGGTVAAQYAALRPDKIDGLLLLAAYPAPNVNLHDARPPALSLLAEKD 219

Query: 173 HILNWEKYEEAKKRLPNSSTFRTIVGGNHSRFGNYGHQKGDGKATLSHKSSRKQLATFIS 232
            + +      +RLP ++    + G  HS FG YG Q+GDG  T+S   +E+++   +
Sbjct: 220 GVADAGLVRGGLERLPKMTRLTVLPGAVHSFPGRYGPQQGDGVPTVSRARAEREIVQAVE 279

Query: 233 NFI 235
           FI
Sbjct: 280 TFI 282
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5504 (GBS158) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 4; MW 27kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 5; MW 52kDa).

The GBS158-GST fusion product was purified (Figure 113; see also Figure 201, lane 4) and used to immunise mice (lane 1+2 product; 14.5µg/mouse). The resulting antiserum was used for Western blot, FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1771

A DNA sequence (GBSx1878) was identified in *S.agalactiae* <SEQ ID 5505> which encodes the amino acid sequence <SEQ ID 5506>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0281(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06539 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 83/197 (42%), Positives = 114/197 (57%), Gaps = 4/197 (2%)

Query: 35  NTYYLVNDQAV-ILIDPGSNGQEIIAKIKSFEKPLVAILLTHTHYDHIFSLDLVRDTFDN 93
           N Y   NDQ   I+ DPG   +++I  ++  +   +AILLTH H+DHI +++ VR+TF +
Sbjct: 14  NWYIQTNDQGEGIIFDPGGEVEKLITWLRDRQITPLAILLTHAHFDHIGAVEDVRNTF-H 72

Query: 94  PPVYVSEKEAAWLSSPDDNLSGLGRHDDIINVIARPAENFFKLKQPYQLNGFEFTVLPTP 153
            PVY+ E E  WL  P  N S L      I    AR AE+     +Q   +  F + VL TP
Sbjct: 73  IPVYIHENEKEWLIDPQRNGSSLFTPGSSIK--AREAEHLITGEQDLSIGSFSYQVLETP 130

Query: 154 GHSWGGVSPVFHSDKLVVTGDALFRETIGRTDLPTSNFEDLITGIRQKLPTLPSHYSVHP 213
           GHS G +S+     D+ V +GDALF  +IGRTDLP  + + L+  I  +L  LP   +V
Sbjct: 131 GHSPGSLSYYAKEDKIVFSGDALFAGSIGRTDLPGGDHQLLLDSIHDKLLELPEDTTVAS 190

Query: 214 GHQMNTTIGHEKNFNPF 230
           GHG  TTIGHE + NPF
Sbjct: 191 GHGPTTTIGHEMDGNPF 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5507> which encodes the amino acid sequence <SEQ ID 5508>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0407(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 217/231 (93%), Positives = 224/231 (96%)

Query: 1   MPFIFRHSFFNKVLIFWYTIIMKIYKTINHIAGENTYYLVNDQAVILIDPGSNGQEIIAK 60
           +PFIFR+SFFNKVLIFWYTI+MKIYKTINHIAGENTYYLVNDQAVILIDPGSNGQEIIAK
Sbjct: 1   LPFIFRYSFFNKVLIFWYTIMKIYKTINHIAGENTYYLVNDQAVILIDPGSNGQEIIAK 60

Query: 61  IKSFEKPLVAILLTHTHYDHIFSLDLVRDTFDNPPVYVSEKEAAWLSSPDDNLSGLGRHD 120
           IKSFEKPLVAILLTHTHYDHIFSLDLVRD FD+PPVYVSEKEAAWLSSPDDNLSGLGRHD
```

```
Sbjct: 61  IKSPEKPLVAILLTHTHYDHIPSLDLVRDAFDHPPVTVSEKEAAWLSSPDDNLSGLGRHD 120

Query: 121 DIINVIARPAENPFKLKQPYQLNGFEFTVLPTPGHSWGGVSFVFHSDELVVTGDALFRET 180
           DII VIARPAENPFKLKQPYQLNGFEFTVLPT GHSWGGVSFVFHSDELVVTGDALFRET
Sbjct: 121 DIITVIARPAENPFKLKQPYQLNGFEFTVLPTSGHSWGGVSFVFHSDELVVTGDALFRET 180

Query: 181 IGRTDLPTSNPEDLITGIRQELFTLPSHYSVHPGHGMNTTIGHEKNFNPFF 231
           IGRTDLPTSNPEDLITGIRQELFTLP+HY V+PGHG +TTI HEKN NPFF
Sbjct: 181 IGRTDLPTSNPEDLITGIRQELFTLPNHYRVYPGHGPSTTICHEKNANPFF 231
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1772

A DNA sequence (GBSx1879) was identified in *S.agalactiae* <SEQ ID 5509> which encodes the amino acid sequence <SEQ ID 5510>. This protein is predicted to be acetoin reductase (fabG). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1596(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9631> which encodes amino acid sequence <SEQ ID 9632> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC48769 GB:U71200 acetoin reductase [Bos taurus]
 Identities = 162/254 (63%), Positives = 188/254 (73%), Gaps = 2/254 (0%)

Query: 12  KVAIVTGAGQGIGFAIAKRLHADGFKIGVLDYNEETAQAAVDKLSPED--AVAVVADVSK 69
           KVA+VTG QGIG AI   L ADGF + V D NE  ++      +       A+AV  DVS
Sbjct: 4   KVAMVTGAQGIGEAIVXXLSADGFAVAVADLNEAKSKVATDIEKNGGTAIAVKLDVSD 63

Query: 70  RDQVFDAFQKVVDTFGDLNVVVNNAGVAPTTFLDTITEQFEKAFAINVGGTIWGSQAAQ 129
           R+  F A ++V +  G  +V+VNNAG+ PTTF+DTIT E F+K + INV G IWG QAA
Sbjct: 64  RESFFAAVKEVAEKLGGFDVLVNNAGLGPTTFIDTITPELFDKVYHINVAGDIWGIQAAV 123

Query: 130 KHFRELGHGGKIINATSQAGCEGNPNLTVYGGTKFAVRGITQTLAKDLASEGITVNAYAP 189
           + F++ G+GGKIINATSQAG  GNPNL++Y  TKFAVR +T   A+DLA + ITVNAYAP
Sbjct: 124 EQFKKNGNGGKIINATSQAGVVGNPNLSLYSSTKFAVRCLTPVAARDLAEQNITVNAYAP 183

Query: 190 GIVKTPMMFDIAHEVGKNAGKDDEWGMEQFAKDITLKRLSEPEDVANAVGFLAGDDSNYI 249
           GIVKTP  FDIAHEVGKNAGKDDEWGM+ FAKDI LKRLSEPEDVA AV FLAG DSNYI
Sbjct: 184 GIVKTPXXFDIAHEVGKNAGKDDEWGMQTFAKDIALKRLSEPEDVAAAVAFLAGPDSNYI 243

Query: 250 TGQTIVVDGGMVPH 263
           TGQTI VDGGM PH
Sbjct: 244 TGQTIHVDGGMQPH 257
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5511> which encodes the amino acid sequence <SEQ ID 5512>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1131(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 209/213 (98%), Positives = 212/213 (99%)

Query: 1   MTKEYEVSDMSKVAIVTGAGQGIGFAIAKRLHADGFKIGVLDYNEETAQAAVDKLSPEDA 60
           +TK+YEVSDMSKVAIVTGAGQGIGFAIAKRLHADGFKIG+LDYNEETAQAAVDKLSPEDA
Sbjct: 1   LTKKYEVSDMSKVAIVTGAGQGIGFAIAKRLHADGFKIGILDYNEETAQAAVDKLSPEDA 60

Query: 61  VAVVADVSKRDQVFDAFQKVVDTFGDLNVVVNNAGVAPTTPLDTTTEEQFEKAFAINVGG 120
           VAVVADVSKRDQVFDAFQKVVDTFGDLNVVVNNAGVAPTTPLDTTTEEQFEKAFAINVGG
Sbjct: 61  VAVVADVSKRDQVFDAFQKVVDTFGDLNVVVNNAGVAPTTPLDTTTKEQFEKAFAINVGG 120

Query: 121 TIWSGQAAQKHFRESLGHGGKIINATSQAGCEGNPNLTVYGGTKFAVRGITQTLAKDLASE 180
           TIWSGQAAQKHFRESLGHGGKIINATSQAGCEGNPNLTVYGGTKFAVRGITQTLAKDLASE
Sbjct: 121 TIWSGQAAQKHFRESLGHGGKIINATSQAGCEGNPNLTVYGGTKFAVRGITQTLAKDLASE 180

Query: 181 GITVNAYAPGIVKTPMMFDIAHEVGKNAGKDDE 213
           GITVNAYAPGIVKTPMMF IAHEVGKNAGKDDE
Sbjct: 181 GITVNAYAPGIVKTPMMFAIAHEVGKNAGKDDE 213
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1773

A DNA sequence (GBSx1880) was identified in *S.agalactiae* <SEQ ID 5513> which encodes the amino acid sequence <SEQ ID 5514>. This protein is predicted to be ATP-dependent DNA helicase. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3735(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB38451 GB:L47709 22.4% identity with Escherichia coli
          DNA-damage inducible protein ...; putative [Bacillus subtilis]
 Identities = 132/461 (28%), Positives = 231/461 (49%), Gaps = 22/461 (4%)

Query: 21  RKYAVVDLEATGAGPNAS--IIQVGIVIIQGNKIIDSYETDVNPHESLDEHIVHLTGITD 78
           +++ V+D+E TG  P      IIQ+  V+I+  +I + +   +NP++S+   I  LTGI++
Sbjct: 4   QRFVVIDVETTGNSPKKGDKIIQIAAVVIENGQITERPSKYINPNKSIPAFIEQLTGISN 63

Query: 79  KQLAKAPDFGQVAHHIYQLIEDCIFVAHNVKFDANLLAEQLFLEGCELRTPRI-DTVELS 137
           + +     F  VA  ++QL++    FVAHN+ FD   +  +L  G +L    + DTVELS
Sbjct: 64  QMVENEQPFEAVAEEVFQLLDGAYFVAHNIHFDLGFVKYELHKAGFQLPDCEVLDTVELS 123

Query: 138 QVFYPCLEKYSLGALAESLNIELTDAHTAIADARATAQLFIKLKAKISSLPKEVLETILT 197
           ++ +P  E Y L  L+E L +     H A +DA  T  +F+++  K+  LP   L+  +
Sbjct: 124 RIVFPGMEGYKLNSLSEELQLRHDQPHRADSDAEVTGLLFIELLEKLRQLPYPTLKQLRR 183

Query: 198 FADNLLFESYLLIEEAYQEADFVNPKEYYFWQSLVLKKEKAVGKPKKLSSDFQ------- 250
            + + + +   L++     E        Y  +    +++ +A+         +F
Sbjct: 184 LSQHFISDLTHLLDMFINENRHTEIPGYTRFSSFSVREPEAIDVRINEDENFSFEISWE 243

Query: 251 ------VNMALLGMDARPKQVVFADLVKAHFNDQTTTPLEAQPGLGKTYGYLLP--LLDQ 302
                ++  + G + R  Q++     V  F ++     +EA PG+GKT GYL+P  L  +
Sbjct: 244 AGNEKALSELMPGYEKRDGQMMMMREVADAFANREHALIEAPEGIGKTIGYLIPAALFAK 303
```

```
Query: 303 SQKQQIIVSVPTKILQDQIMAKEIKHIQELFHIPCHS--IKGPRNYLKLDAFYKSLQVQD 360
           K+ +I+S  + +LQ QI+ K++  +Q+LF  P  +  +KG  +YL L  F + L  +D
Sbjct: 304 KSKKPVIISTYSTSLLQQQILTKDLPIVQDLFPFFVTAAILKQGSHYLCLYKFSQVLHEED 363

Query: 361 RNRLINRFFMQLLVWLTETTTGDLDEIKQKQRLESYFDQLKHDGK-VTQSSLFYDLDFWK 419
            N      K QLLVWLTET TGD+ E+      +  +D+L +D +   +S   + + F++
Sbjct: 364 DNYDAVLTKAQLLVWLTETNTGDVAELNLPSGGKLLWDRLAYDDDSYRSRSEHVIGFYS 423

Query: 420 RSYDKVAQSQLVIINHAYYL-ERVQDDKDFARGKVLVFDEA 459
            R+    +S LVI NH+  L +     K    + DEA
Sbjct: 424 RAKQLMRSRLKVITNHSLLLTDEGSHKKRLPESGTFIIDEA 464
 Identities = 63/195 (32%), Positives = 88/195 (44%), Gaps = 16/195 (8%)

Query: 629 KVWIDTSMPNILDLSPSQYAYKIAKRLQDIMTLKQPT-LVLITSKQTMPMVSDYLDKWEI 687
            +V I   M +I D    ++ + A+ ++ +   KQP LVLITSKQTMPMVSDYLDKWEI
Sbjct: 720 QVMIPKEMKSIQDTGQPEFIQDTARYIELMAKEKQPKIIVLFTSHDMLKKVHQ-----EL 774

Query: 688 KH---------LTQD-KNGLAYNVKKRPDRGESNLLLGTGSFWEGVDFVHRDRLIEVITR 737
            KH         L Q    G    + K F      +LLGT  FWEGVDF    + I E R
Sbjct: 775 KHNMSASGIQLLAQGITGSSPGKLMKTFKTSNQALLLGTNHFWEGVDFPGDSLITVMIVR 834

Query: 738 LPFDTPKDYFIQKLSQSLTKEGKNFFYDYSLIMTVLKLKQALQRTTRREEQKSAVILLDS 797
            LPF +P        +  K+GKN F   SLP  VL  +Q +GR  R    K  +IILD
Sbjct: 835 LPFRSPDHPLHAAKCKLARKKGKNPFQTVSLPEAVLTFRQSIGRLLRSAGDKGTIILLDR 894

Query: 798 RLVIKSYGQTIMHSL 812
            R+    YG+  + +L
Sbjct: 895 RIKTAGYGRLFLDAL 909
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5515> which encodes the amino acid sequence <SEQ ID 5516>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3735(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 500/835 (59%), Positives = 626/835 (74%), Gaps = 2/835 (0%)

Query: 1   MFCPIDIACYNRLTMTQKKLRKYAVVDLEATGAPNASIIQVGIVIIQGNKIDSYETDV 60
           MFCPIDIACYNRLTMTQKKLRKYAVVDLEATGAPNASIIQVGIVIIQGNKIDSYETDV
Sbjct: 1   MFCPIDIACYNRLTMTQKKLRKYAVVDLEATGAGPNASIIQVGIVIIQGNKIDSYETDV 60

Query: 61  NPHESLDSHIVHLTGITDKQLAKAPDFGQVAHHIYQLIEDCIFVAHNVKFDANLIAEQLF 120
           NPHESLDSHIVHLTGITDKQLAKAPDFGQVAHHIYQLIEDCIFVAHNVKFDANLIAE LF
Sbjct: 61  NPHESLDSHIVHLTGITDKQLAKADPDFGQVAHHIYQLIEDCIFVAHNVKFDANLIAEALF 120

Query: 121 LEGCRKLRTPRIDTVELSQVFYPCLRKYSLGALARSLNIKLDTDAHTAIADARATAQLFIKL 180
           LEG EL  PR+DTVELQ+F+P  EKY+L  L+  LNI+L  +AHTAIADARATA LF++L
Sbjct: 121 LEGYELTIPRVDTVELAQLFFPREFKYNLSHLSRQLNIDLAEARTAIADARATATALFLRL 180

Query: 181 KAKISSLDKEVLETLLTFADNLLPESYLLIERAYQEADFVNPKEYYFWQGLVLKKEKAVG 240
            KI SLP E LE++L ++D+LLPE+ ++I+E   +A   +P +Y   + ++L K
Sbjct: 181 LGKISSLPIECLESLAVYSDSSLPFTAMVIQEGLAKAKFYDPMKYIKIRQILPRGSKAL 240

Query: 241 KPKKLSSDFQVNMALLGMDARPKQVVFADLVKAHFNDQTTTFLEAQPGLGKTYGYLLPLL 300
           KP ++S  F +NMALLG++ RPKQ  FA L+   ++  +F+EAQ G+GKTYGYLLPLL
Sbjct: 241 KFYQISKSFPINMALLGLRRRPKQTQFAQLIDRDYHQSVASFIEAQTGIGKTYGYLLPLL 300

Query: 301 DQSQKQQIIVSVPTKILQDQIMAKEIKHIQELFHIPCHSIKGPRNYLKLDAFYKSLQVQD 360
               +  + QIIVSVPTK+LQDQ+MA E+  IQE FHI  CHS+KGP NYLKLD+F  SL   D
```

```
Sbjct: 301 AKEDQNQIIVSVPTKLLQDQLMAGEVAAIQEQPHIACHSLKGPANYLKLDSFADSLDQND 360

Query: 361 RNRLINRPFKMQLLVWLTETTTGDLDEIKQKQRLESYFDQLKHDGEVTQSSLFYDLDFWKR 420
            +NRL+NR+KMQLLVWL ET TGDLDEIKQKQR  +YF+QLKHDG++ QSS FYD DFW+
Sbjct: 361 QNRLVNRYKMQLLVWLLETKTGDLDEIKQKQRFAAYFEQLKHDGDIKQSSSKFYDYDFWRV 420

Query: 421 SYDKVAQSQLVIINHAYFLERVQDDKDFANGKVLVFDEAQKLVLGLENVFSRCQLDISHQL 480
            SY+K   ++L+I NEAYFL RVQDDKDFA+ KVLVFDEAQKL+L L+  SR QL+++  L
Sbjct: 421 SYEKAKTARLLIITNHAYFLHRVQDDKDFARNKVLVFDEAQKLMLQLDQLSRHQLNLTVFL 480

Query: 481 QVIQKIIDSSIPLLQKRLLESISYELSHAVKLFYRHNSFESETWLKRLKNSINALEVVG 540
            Q IQ  + + +PLL+KRLLES+S+EL      +Y++   + +  W  R+      L
Sbjct: 481 QTIQAKLSNPLPLLEKRLLESLSFELGQVSSDYYQNKEHQLAHDW-SRIAGYAKELTGAD 539

Query: 541 LLELQTFFTATTYNYWFKTDKVNKKRLITLRGRAREDPLKFSKFLPPTKKTYMISATLQIS 600
              ELQ FF  +  +YW  ++K  EKR+T L  A + F+ F + LP T KTY +SATL IS
Sbjct: 540 YQELQAFFATSDGDYWLSSKKQEEKRYTYLNSASKAFIHPQQLLPETVKTYFVSATLTIS 599

Query: 601 PKVVLSDLLGGFSSISTEKIAHEKNANQKVWIDTSMPNIILDLSPEQYAYETAKRLQDIMT 660
             +V L+DLL GF      I  +K  +Q V +D   P + ++S + Y   IAKR++ +
Sbjct: 600 SEVTLADLL-GFEEYLYHVIEKDKKDQQLVLVDQSAPIVTEVSDQIYVEAIAKRIESLKQ 658

Query: 661 LKQPTLVLLTSKQTMPMVSDYLDKWEIKHLTQDKNGLAYNVKKRFDRGESNLLLGTGSFW 720
             P LVL  SK+ + +VSDYLD+W++ HL Q+KNG AYN+KKRFD+GE   +LLG GSFW
Sbjct: 659 EGYPILVLFNSKKHLLLVSDYLDQWVPHLAQEKNGTAYNIKKRFDQGEQTILLGLGSFW 718

Query: 721 EGVDFVHRDRLIEVITRLPFDTPKDYFIQKLSQSLTKEGKNFFYDYSLPMTVLKLKQALS 780
            EGVDF+  DR+I +I RLPFD P+D+F++K+S  L ++GKN F DY LPMT+L+LKQA+G
Sbjct: 719 EGVDFIQADRMITLIARLPFDNPKDFFVKKMSHYLLEKGKNPFRDYFLPMTILRLKQAIG 778

Query: 781 RTTRREQQKSAVIILDSRLVIKSYQQTIMHSLGRDFEISKEKINKVLTEWAKFLI 835
            RT RR++QKS VIILD RL+ KSYQQ I+  LG++F IS++  +  L E   FLI
Sbjct: 779 RTNRRQDQKSVVIILDRRLLTKSYQQVILEGLQEFLISQQNFHDCLVETDCFLI 833
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1774

A DNA sequence (GBSx1881) was identified in *S.agalactiae* <SEQ ID 5517> which encodes the amino acid sequence <SEQ ID 5518>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2042(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9633> which encodes amino acid sequence <SEQ ID 9634> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF12702 GB:AF035157 aspartate aminotransferase [Lactococcus
             lactis]
 Identities = 270/391 (69%), Positives = 314/391 (80%)

Query: 7   MTYLSERVLNMEESVTLAAGAKARKLRVQGRDILSLTLGEPDFATPKNIQQAAIRAITDG 66
            M   S+ VL M+ESVTLAA  +A+ L+ QGRDI+ LVTLG+PDF TPK I QAAIEAI +G
Sbjct: 1   MKKCSDFVLKMDESVTLAAANRAKALKAQGRDIIDLVTLGQPDFFTPKKIGQAAIEAINNG 60

Query: 67  RASFYTPSSGLPELKSAINAYFERFYGYSLKPNQVVVGTGAKFILYTFFMTVLNPGDEVI 126
            +ASFYT + GLPELK A+  Y+ RFY Y ++ N++++  GAKF LY +FM  ++P DEVI
```

```
Sbjct: 61  QASFYTQAGGLPELKKAVQHYWTRFYAYEIQTNSLITAGAKFALYAYPMATVDPLDEVI 120

Query: 127 IPTFYWVSYADQIKMAEGKPVPVTAKEVNHFKVTVEQLEAVRTDKTKVILLNSPSNPTGM 186
            IP FYWVSY DQ+KMA G PV V AK+ N+FKVTVEQLE  RT KTK++LLNSPSNPTGM
Sbjct: 121 IPAFYWVSYVDQVKMAGGNPVIVEAKQENNFKVTVEQLEKARTSKTKILLLNSPSNPTGM 180

Query: 187 IYKAEELEAIGNWAVEHDILILADDIYGRLVYNGNIPTPISSLSESIRNQTIVINGVSKT 246
            IY  EEL AIG WAV HD+LLIADDIY RLVYNG  PT ISSLS+ IRN+T VINGVSKT
Sbjct: 181 IYSKEELTAIGEWAVAHDLLIIADDIYHRLVYNGAEPTAISSLSDEIRNRTTVINGVSKT 240

Query: 247 YAMTGWRVGFAVGKEDIIAAMSKVVSQTTSNLTAVSQYATIEALNGSQESFEKMRLAFEE 306
            +AMTGWR+G AVG+ +IIAAM+K+ SQTTSN TAV+QYA IEA   + +SFEKM  AFEE
Sbjct: 241 FAMTGWRIGLAVGDPEIIAAMTKIASQTTSNPTAVAQYAAIEAFEENDKSFEKMHAAFER 300

Query: 307 RLNIIYPLLCQVPGFEVVKPQGAFYLFPNVTKAMEMKGYTDVTAFTDAILEEVGLALVTG 366
            RLN IY  L +VPGFE+VKP GAFYLFP VTKAM MKGYTDVT FT AILEE G+ALVTG
Sbjct: 301 RLNKIYLQLSEVPGFKLVKPNGAFYLFPKVTKAMAMKGYTDVTDFTTAILEEAGVALVTG 360

Query: 367 AGFGAPENVRLSYATDLETLKEAVRRLHVFM 397
            AGFG+PENVRLSYAT LETL+ AV RL  +M
Sbjct: 361 AGFGSPENVRLSYATSLETLEAAVIRLKDWM 391
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1005> which encodes the amino acid sequence <SEQ ID 1006>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.48   Transmembrane   95 - 111 (  95 - 113)

----- Final Results -----
               bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 301/397 (75%), Positives = 343/397 (85%)

Query: 7   MTYLSERVLNMEESVTLAAGAKARELRVQGRDILSLTLGEPDFATPKNIQQAAIEAITDS 66
            M  LS+RVL M+ESVTLAAGA+A+ L+ QGRD+L+LTLGEPDF TPK+IQ  AIE+I +G
Sbjct: 1   MPKLSKRVLEMKESVTLAAGARAKALKAQGRDVLNLTLGEPDFPTPKHIQDKAIESIQNG 60

Query: 67  RASFYTPSGGLPELKSAINAYFERFYGYSLKPNQVVVGTGAKFILYTFPMTVLNPGDEVI 126
             ASFYT +SGLPELK+AI  Y +  YGY L P+Q+V GTGAKFILY FPM VLNPGD+V+
Sbjct: 61  TASFYTNASGLPELKAAIATYLKNQYGYHLSPDQIVAGTGAKFILYAFPMAVLNPGDQVL 120

Query: 127 IPTFYWVSYADQIKMAEGKPVPVTAKEVNHFKVTVEQLEAVRTDKTKVILLNSPSNPTGM 186
            IPTFYWVSY+DQ+KMAEG+P+PV   E N FKVTV+QLE  RT KTKV+L+NSPSNPTGM
Sbjct: 121 IPTFYWVSYSDQVKMAEGQPIPVQGLERNQFKVTVDQLERARTSKTKVVLINSPSNPTGM 180

Query: 187 IYKAEELKAIGNWAVEHDILILADDIYGRLVYNGNIPTPISSLSESIRNQTIVINGVSKT 246
            IY AEEL AIG WAV +DILILADDIYG LVYNGN F PIS+LSE+IR QTI +NGV+K+
Sbjct: 181 IYGAEELKAIGEWAVHNDILILADDIYGSLVYNGNQFVPISTLSEAIRRQTITVNGVAKS 240

Query: 247 YAMTGWRVGFAVGNHDIIAAMSKVVSQTTSNLTAVSQYATIEALNGSQESFEKMRLAFEE 306
            YAMTGWRVGFA G  +II+AMSK++ QTTSNLT VSQYA IEA  GSQ S E+MRLAFEE
Sbjct: 241 YAMTGWRVGFAAGEPEIISAMSKIIGQTTSNLTTVSQYAAIEAFCGSQQSSLEEMRLAFEE 300

Query: 307 RLNIIYPLLCQVPGFEVVKPQGAFYLFPNVTKAMEMKGYTDVTAFTDAILEEVGLALVTG 366
            RLNI YPLLCQVPGFEVVKPQGAFY FPNV KAMEM G+DVT+F +AILEEVGLA+V+G
Sbjct: 301 RLNITYPLLCQVPGFEVVKPQGAFYFFPNVKKAMEMIGFSDVTSFANAILEEVGLAVVSG 360

Query: 367 AGFGAPENVRLSYATDLETLKEAVRRLHVFMGSNRIN 403
            AGFGAPENVRLSYATD+ETLKEAVRRLHVFM SNRIN
Sbjct: 361 AGFGAPENVRLSYATDIETLKEAVRRLHVFMESNRIN 397
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1775

A DNA sequence (GBSx1882) was identified in *S.agalactiae* <SEQ ID 5519> which encodes the amino acid sequence <SEQ ID 5520>. This protein is predicted to be asparaginyl-tRNA synthetase (asnS). Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1488(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05415 GB:AP001512 asparaginyl-tRNA synthetase [Bacillus halodurans]
 Identities = 252/442 (57%), Positives = 316/442 (71%), Gaps = 15/442 (3%)

Query: 7   SIVDVKDYVGQEVTIGAWVANKGGKGKIAFVQLRDGSAFPQGVAFKPNFIEKYGEESGLE 66
            +I  +  YV QEVT+GAW+ANK   GKIAF+QLRDG+ F QGV  K          E G E
Sbjct: 4   TIAKIGQYVEQEVTLGAWLANKRSSGKIAFLQLRDGTGFIQGVVVKA--------EVGDE 55

Query: 67  KFDVIKRLNQETSVVVTGIVKEDERSKPGYELDITDLEVIGESHEYPITPKEHGTDFLMD 126
             F   K L QE+S+YVTGIV++DER+  GYEL +T  ++I E+ +YPITPKEHGT+FLMD
Sbjct: 56  WFQKAKNLTQESSLYVTGIVRKDERAPSGYELTVTSFDIIHEATDYPITPKEHGTEFLMD 115

Query: 127 NRHLWLRSRKQMAVMQIRNAIIYSTYEFFDQNGFIKFDSPILSENAAEDSTELFETDYFG 186
            +RHLM+RSRKQ AV++IRN II +TYEFF +NGF+K D PIL+ +A E +TELF T YF
Sbjct: 116 HRHLWIRSRKQHAVLRIRNSIIRATYEFFHENGFVKVDPPILTGSAPEGTTELFHTKYFD 175

Query: 187 KPAFLSQSGQLYLEAGAMALGRVFDFGPVFRAEKSKTRHLTEFWMMDAEYSFLSHEESI 246
            + AFLSQSGQLY+EA A+A GRVF PGP FRAEKSKTRHL EFWM++ E +F+  EESL
Sbjct: 176 EDAFLSQSGQLYMEAAALAFGRVFSFGPTFRAEKSKTRHLIEFWMIEFPMAFVEFEESL 235

Query: 247 DLQEAYVKALIQGVLDRAPQALDILERDVEALKRYIAEPFKRVSYDDAITLLQKHEADED 306
            ++QE YV  ++Q VL      L  L RD   L+  I  PF R+SYDDAI  L E   D+
Sbjct: 236 EIQSNYVAYIVQSVLKHCAIELKTLGRDTSVLRS-IQAPFPRISYDDAIKFLHEKGFDD- 293

Query: 307 TDYEHLEWGDDFGSPHETWISNYFGVPTFVVNYPASPKAFYMKPVPGNPERVLCADLLAP 366
              +E GDDFG+PHET I+  +F  P F+  +YP S K FYM+P P   + VLCADL+AP
Sbjct: 294 -----IEWGDDFGAPHETAIAEHFDKPVFITHYPTSLKPFYMEPDPNRDDVVLCADLIAP 348

Query: 367 EGYGEIIGGSMREDDYDALVAKMDELGMDKSEYDFYLDLRKYGSVPHGGFGIGIERMVTF 426
            EGYGEIIGGS R  DYD L  +++E  +     Y +YLDLRKYGSVPH GFG+GIERVV +
Sbjct: 349 EGYGEIIGGSQRISDYDLLKKRLEEHDLSLDAYAWYLDLRKYGSVPHSGFGLGLERTVGW 408

Query: 427 VAGTKHIREAIPFPRMLHRIKP 448
            ++G  H+RE IPFPR+L+R+ P
Sbjct: 409 ISGAGHVRETIPFPRLLNRLYP 430
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5521> which encodes the amino acid sequence <SEQ ID 5522>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1488(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 443/448 (98%), Positives = 447/448 (98%)

Query: 1   MSKKLISIVDVKDYVGQEVTIGAWVANKSGKGKIAFVQLRDGSAFFQGVAFKPNFIRKYG 60
           MSKKLISIVDVKDYVGQEVTIGAWVANKSGKGKIAFVQLRDGSAFFQGVAFKPNFIRKYG
Sbjct: 1   MSKKLISIVDVKDYVGQEVTIGAWVANKSGKGKIAFVQLRDGSAFFQGVAFKPNFIRKYG 60

Query: 61  KESGLSKPDVIKRLNQETSVYYTGIVKKDERSKPGYELDITDLRVIGESHEYPITPKEHG 120
           EESGLSKPDVIKRLNQETSVYYTGIVKKDERSKPGYELDITDLR+IGESHEYPITPKEHG
Sbjct: 61  EESGLSKPDVIKRLNQETSVYYTGIVKKDERSKPGYELDITDLRIIGESHEYPITPKEHG 120

Query: 121 TDFLMDNRHLWLRSRKQMAVMQIRNAIIYSTYEPFDQNGFIKFDSPILSENAAEDSTELF 180
           TDFLMDNRILWLRSRKQMAVMQIRNAIIY+TYEPFDQNGFIKFDSPILSENAAEDSTELF
Sbjct: 121 TDFLMDNRHLWLRSRKQMAVMQIRNAIIYATYEPFDQNGFIKFDSPILSENAAEDSTELF 180

Query: 181 ETDYFGKPAFLSQSGQLYLEAGAMALGRVFDPGPVFRAEKSKTRRHLTEFWMMDAEYSFL 240
           ETDYFGKPAFLSQSGQLYLEAGAMALGRVFDPGPVFRAEKSKTRRHLTEFWMMDAEYSFL
Sbjct: 181 ETDYFGKPAFLSQSGQLYLEAGAMALGRVFDPGPVFRAEKSKTRRHLTEFWMMDAEYSFL 240

Query: 241 SHEESLDLQEAYVKALIQGVLDRAPQALDILERDVEALKRYIAEPFKRVSYDDAITLLQE 300
           SHEESLDLQEAYVKALIQGVLDRAPQALDILERDVEALKRYI EPFKRVSYDDAITLLQE
Sbjct: 241 SHEESLDLQEAYVKALIQGVLDRAPQALDILERDVEALKRYITEPFKRVSYDDAITLLQE 300

Query: 301 HEADEDTDYEHLEHGDDFGSPHETWISNYFGVPTFVVNYPASFKAFYMKPVPGNPERVLC 360
           HEADEDTDYEHLEHGDDFGSPHETWISNYFGVPTFVVNYPASFKAFYMKPVPGNPERVLC
Sbjct: 301 HEADEDTDYEHLEHGDDFGSPHETWISNYFGVPTFVVNYPASFKAFYMKPVPGNPERVLC 360

Query: 361 ADLLAPEGYGEIIGGSMREDDYDALVAKMDELGMDKSEYDFYLDLRKYGSVPHGGFGIGI 420
           ADLLAPEGYGEIIGGSMREDD+DALVAKMDELGMDKSEYDFYLDLRKYGSVPHGGFGIGI
Sbjct: 361 ADLLAPEGYGEIIGGSMREDDNDALVAKMDELGMDKSEYDFYLDLRKYGSVPHGGFGIGI 420

Query: 421 ERMVTFVAGTKHIREAIPFPRMLHRIKP 448
           ERMVTFVAGTKHIREAIPFPRMLHRI+P
Sbjct: 421 ERMVTFVAGTKHIREAIPFPRMLHRIRP 448
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1776

A DNA sequence (GBSx1883) was identified in *S.agalactiae* <SEQ ID 5523> which encodes the amino acid sequence <SEQ ID 5524>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL     Likelihood = -6.85   Transmembrane  103 - 119 ( 102 - 127)
    INTEGRAL     Likelihood = -5.04   Transmembrane   73 -  89 (  68 -  93)
    INTEGRAL     Likelihood = -4.19   Transmembrane   31 -  47 (  31 -  49)
    INTEGRAL     Likelihood = -1.86   Transmembrane  157 - 173 ( 157 - 173)

----- Final Results -----
                bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD40355 GB:AF036485 hypothetical protein [Plasmid pNZ4000]
 Identities = 39/135 (28%), Positives = 72/135 (52%), Gaps = 4/135 (2%)

Query: 3   KSPARLISFISIAIAINLWGANLALFLRLPTYLDTIGTLLIAVILGPWYAASTAFLSALI 62
           K  A  ++ I  A+ IN V   LA  L+LP++L ++GT L +++ GP   A + F++ +I
Sbjct: 15  KLSAATMTLIPAAVGINYVAKALARGLKLPWWLGSLGTFLASMLAGPVAGAISGFINNVI 74

Query: 63  NWMTTDIFSLYYSPVAIVVAIITGILIKRNCKPSS--LLWKSLIISLPGTIIASVITVIL 120
```

```
              +T    S  Y+  +I + I  G+L       S+  +    ++II++    +I++ + VI
Sbjct: 75  YGLTLSPISTVYAITSIGIGIAVGVLHANGWFSSARRVFVSAIIIAIVSAVISTPLNVIF 134

Query: 121 FKGIT--SSGSSIIA 133
           + G T  + G S+ A
Sbjct: 135 WGGQTGIAWGDSLFA 149
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1777

A DNA sequence (GBSx1884) was identified in *S.agalactiae* <SEQ ID 5525> which encodes the amino acid sequence <SEQ ID 5526>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1873(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC75223 GB:AE000305 orf, hypothetical protein [Escherichia coli K12]
 Identities = 97/305 (31%), Positives = 160/305 (51%), Gaps = 10/305 (3%)

Query: 1   MNKEKIIIDCDPGIDDTLALMYAIQHPKLEVVAITTTAGNSPVELGLKNTFVTLELLNRH 60
           M K KII+DCDPG DD +A+M A +HP ++++ ITI AGN  ++  L N      + L
Sbjct: 1   MEKRKIILDCDPGHDDAIAIMMAAKFPAIDLIGITIVAGNQTLDKTLINGLNVCQKL-EI 59

Query: 61  DIPVYVGDNLPLQREFVSAQDTHGMDGLGENNFTLAQPIIPQKESADC---FLANYFEHK 117
           ++PVY G   P+ R+ + A + HG  GL    F   +P+  Q ES       +
Sbjct: 60  NVPVYAGMPQPINRQQIVADNIHGETGLDGPVF---EPLTRQAESTHAVKYIIDTLMASD 116

Query: 118 NDTSIIALGPLTNIARALQTNPKLGKHCKRFISMGSFKSHGNCSPVAEYNYWCDPHAAQ 177
            D ++ +GPL+NIA A++  P +    + M3G++ + GN +P AR+N + DP AA+
Sbjct: 117 GDITLVPVGPLSNIAVAMRMQPAILPKIREIVLMGGAYGT-GNFTPSAEFNIFADPEAAR 175

Query: 178 YVFENLDKKIEMVGLDITRHIVLTPNHLSYMERINPDVSSFIQKITKFYFDFHWQYEHII 237
            VF +    + M+GLD+T   V TP+ ++ MER           I  F    ++    +
Sbjct: 176 VVFTS-GVPLVMMGLDLTNQTVCTPDVIARMSRAGSPAGELFSDIMNFTLKTQFENYGLA 234

Query: 238 GCVINDPLAIAYFVNENIATGFDSYTDVACH-GIAMGQTIVDQYHFYKKDANSKILTSVN 296
            G  ++D   I Y +N +    + Y +V  + G   G+T+ D+    K AN+K+  +++
Sbjct: 235 GGPVHDATCIGYLINPDGIKTQEMYVEVDVNSGPCYGRTVCDELGVLGKPANTKVGITID 294

Query: 297 TNLFW 301
           T+ FW
Sbjct: 295 TDWFW 299
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1778

A DNA sequence (GBSx1885) was identified in *S.agalactiae* <SEQ ID 5527> which encodes the amino acid sequence <SEQ ID 5528>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1860(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB62728 GB:AL133423 hypothetical protein SC4A7.24c
            [Streptomyces coelicolor A3(2)]
 Identities = 36/134 (26%), Positives = 57/134 (41%), Gaps = 7/134 (5%)

Query: 1   MLYEVTSSNTQGVDGKVYLSNGKIVETNNPLWHL----PGFNDEELIALAWSTCLNATIK 56
            +LY   ++    G DG+V   +G++    +P    +      G NPE+L A  +S C    +
Sbjct: 8   VLYTAVATAENGRDGRVATDDGRLDVVVNPPKEWGGNGAGTNPEQLFAAGYSACFQGALG 67

Query: 57  AILEQKGFKDLKSRVDVTCQFMKEKQVGKGFYFQVNAVASIEKLSLSDSKLIVNKAHSRC 116
            +  Q+G     S V     + K      GF    V    A  I   +   + +V KAH  C
Sbjct: 68  VVARQEGADISGSTVTAKVGIGKNDD---GFGIIVEISAEIPTVDAATARSLVEKAHQVC 124

Query: 117 PISKLISNAKTINL 130
            P SK      T+ L
Sbjct: 125 PYSKATRGNITVTL 138
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1779

A DNA sequence (GBSx1886) was identified in *S.agalactiae* <SEQ ID 5529> which encodes the amino acid sequence <SEQ ID 5530>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0531(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9635> which encodes amino acid sequence <SEQ ID 9636> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15482 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 164/285 (57%), Positives = 207/285 (72%), Gaps = 2/285 (0%)

Query: 6   IKLVIVTGMSGAGKTVAIQSFEDLGYFTIDNMPPTLVPKFLELAAQSGDT-SKIAMVVDM 64
            I+LVI+TGMSGAGKTVAIQSFEDLGYF +DN+PP+L+PKFLEL  +S     SK+A+V+D+
Sbjct: 9   IQLVIITGMSGAGKTVAIQSFEDLGYFCVDNLPPSLLPKFLELMKESNSKMSKVALVMDL 68

Query: 65  RSRDLFFREINSILDSLEINDNINFKILFLDATDTELVSRYKETRRSHPLAADGRVLDGIS 124
            R R FF  +   LD +  N  I  +ILFLDA D+  LV+RYKETRRSHPLAA G  L+GI+
Sbjct: 69  RGREFFDRLIEALDEMAENPWITPRILFLDAKDSILVTRYKETRRSHPLAATGLPLEGIA 128

Query: 125 LERELLAPLKSMSQNVVDTSELTPRQLRKVISKEPSNQQSQSSPRIEVMSFGFKYGIPLD 184
            LERELL LK  SQ + DTS++ PR LR+ I K F+      ++ F + VMSFGFKYGIP+D
Sbjct: 129 LERSLLEELKGRSQIIYDTSDMKPRDLREKIVKHFATNQGET-FTVNVMSFGFKYGIPID 187

Query: 185 ADLVFDVRFLFNPYYKPELRDKTGLDTEVYDTVMSFDESDDFYDHLLALIKPILPGYQNE 244
```

```
              ADLVFDVRFLPNFYY   +R  TG D EV  YVM ++E+  F + L+ L+  +LP Y+ E
Sbjct: 188 ADLVFDVRFLPNFYYIESMRPLYGKDKEVSSYVMKWNETQKFNEKLIDLLSFMLPSYKRE 247

Query: 245 GKSVLTVAIGCTGGQHRSTAFAERLSEDLKADWTVNESHRDKNKR 289
           GKS + +AIGCTGGQHRS   A  L++  K D+  + +HRD  KR
Sbjct: 248 GKSQVVIAIGCTGGQHRSVTLAENLADYFKKDYYTHVTHRDIEKR 292
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5531> which encodes the amino acid sequence <SEQ ID 5532>. Analysis of this protein sequence reveals the following:

```
      Possible site: 20
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15482 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 164/291 (56%), Positives = 213/291 (72%), Gaps = 3/291 (1%)

Query: 1   MSDKH-INLVIVTGMSGAGKTVAIQSFEDLGYFTIDNMPPALVPKFLELLEQTNENR-RV 58
           +S+ H I LVI+TGMSGAGKTVAIQSFEDLGYF +DN+PP+L+PKFLEL++++N    +V
Sbjct: 3   VSESHDIQLVIITGMSGAGKTVAIQSFEDLGYFCVDNLPPSLLPKFLELMKESNSKMSKV 62

Query: 59  ALVVDMRSRLFPKEINSTLDSIESNPSIDFRILFLDATDGELVSRYKETRRSHPLAADGR 118
           ALV+D+R R FF  +   LD +  NP I  RILFLDA D  LV+RYKETRRSHPLAA G
Sbjct: 63  ALVMDLRGREFFDRLIEALDEMAENPWITPRILFLDAKDSILVTRYKETRRSHPLAATGL 122

Query: 119 VLDGIRLERELLSPLKSMSQEVVDTTKLTPRQLRKTISDQFSEGSNQASFRIEVMSFGFK 178
            L+GI LERELL   LK  SQ + DT+ + PR LR+ I   F+      + +F + VMSFGFK
Sbjct: 123 PLEGIALERELLEELKGRSQIIYDTSDMKPRDLREKIVKHFATNQGR-TFTVNVMSFGFK 181

Query: 179 YGLPLDADLVFDVRFLPNPYYQVELREKTGLDEDVFNYVMSHPESEVFYKHLLNLIVPIL 238
           YG+P+DADLVFDVRFLPNPYY   +R  TG D++V +YVM   E++ F + L++L+  +L
Sbjct: 182 YGIPIDADLVFDVRFLPNPYYIESMRPLTGKDKEVSSYVMKWNETQKFNEKLIDLLSFML 241

Query: 239 PAYQKEGKSVLTVAIGCTGGQHRSVAFAHCLAESLATDWSVNESHRDQNRR 289
           P+Y++EGKS + +AIGCTGGQHRSV  A  LA+     D+  + +HRD  +R
Sbjct: 242 PSYKREGKSQVVIAIGCTGGQHRSVTLAENLADYFKKDYYTHVTHRDIEKR 292
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 234/296 (79%), Positives = 263/296 (88%)

Query: 1   MSDEQIKLVIVTGMSGAGKTVAIQSFEDLGYFTIDNMPPTLVPKFLELAAQSGDTSKIAM 60
           MSD+ I LVIVTGMSGAGKTVAIQSFEDLGYFTIDNMPP LVPKFLEL  Q+ +  ++A+
Sbjct: 1   MSDKHINLVIVTGMSGAGKTVAIQSFEDLGYFTIDNMPPALVPKFLELIEQTNENRRVAL 60

Query: 61  VVDMRSRLFPREINSILDSLEINDNINFKLFLDATDTELVSRYKETRRSHPLAADGRVL 120
           VVDMRSRLFP+EINS LDS+E N +I+F+ILFLDATD ELVSRYKETRRSHPLAADGRVL
Sbjct: 61  VVDMRSRLFPKEINSTLDSIESNPSIDFRILFLDATDGELVSRYKETRRSHPLAADGRVL 120

Query: 121 DGISLERELLAPLKSMSQNVVDTSELTPRQLRKVISKEFSNQDSQSSSFRIEVMSFGFKYG 180
           DGI LERELL+PLKSMSQ+VVDT++LTPRQLRK IS +FS   +Q+SFRIEVMSFGFKYG
Sbjct: 121 DGIRLERELLSPLKSMSQHVVDTTKLTPRQLRKTISDQFSEGSNQASFRIEVMSFGFKYG 180

Query: 181 LPLDADLVFDVRFLPNPYYKPRLRDKTGLDTEVYDVMSHPESDDFYDHLLALIKPILPG 240
           +PLDADLVFDVRFLPNPYY+ ELR+KTGLD +V++YVMS  ES+ FY HLL LI PILP
Sbjct: 181 LPLDADLVFDVRFLPNPYYQVELREKTGLDEDVFNYVMSHPESEVFYKHLLNLIVPILPA 240

Query: 241 YCNEGKSVLTVAIGCTGGQHRSTAFAERLSRDLKADWTVNESHRDKNRRKETVNRS 296
           Y  EGKSVLTVAIGCTGGQHRS AFAE L+E L  DW+VNESHRD+N+RKETVNRS
Sbjct: 241 YQKEGKSVLTVAIGCTGGQHRSVAFAHCLAESLATDWSVNESHRDQNRRKETVNRS 296
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1780

A DNA sequence (GBSx1887) was identified in *S.agalactiae* <SEQ ID 5533> which encodes the amino acid sequence <SEQ ID 5534>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB96620 GB:AJ400630 hypothetical protein [Streptococcus pneumoniae bacteriophage MM1]
 Identities = 254/321 (79%), Positives = 286/321 (88%), Gaps = 1/321 (0%)

Query: 1   MRKPKITVIGGGTGIPVILKSLRLEDVEITAVVTVADDGGSSGELRSVMQ-LTPPGDLRN 59
            MRKPKITVIGGGTGIPVILKSLR +DVEI A+VTVADDGGSSGELR  MQ LTPPGDLRN
Sbjct: 1   MRKPKITVIGGGTGIPVILKSLREKDVEIAAIVTVADDGGSSGELRNMQQLTPPGDLRN 60

Query: 60  VLVALSDMPKFYEQIPQYRFAEGDGDFAGHPLGNLIIAGVAEMQGSTYNAMQSLTQPFHT 119
            VLVA+SDMPKFYE++PQYRF+E  G FAGHPLGNLIIAG++EMQGSTYNAMQ L++PFHT
Sbjct: 61  VLVANSDMPKFYEKVPQYRFSEDAGAFAGHPLGNLIIAGLSEMQGSTYNAMQLLSKPFHT 120

Query: 120 TGKIYPSSEHPLTLHAVFKDGHEVVGESQIADYKGMIDNVVVTNTYNEETPTASRKVVDA 179
            TGKIYPSS+HPLTLHAVF+DG EV GES I D++G+ID+VVVTN  N++ TP ASR+VV
Sbjct: 121 TGKIYPSSDHPLTLHAVFQDGTEVAGESHIVDHRGIIDNVVVTNALNDDTPLASRRVVQT 180

Query: 180 ILESDMIVLGPGSLFTSILPNLVIPEIKQALLETRAEVAYVCNIMTQRGETEHFTDADHV 239
            ILESDMIVLGPGSLFTSILPN+VI EI +ALLET+AE+AYVCNIMTQRGETEHFTD+DHV
Sbjct: 181 ILESDMIVLGPGSLFTSILPNIVIKEIGRALLETKAEIAYVCNIMTQRGETEHFTDSDHV 240

Query: 240 EVLKRHLGQDAIDTVLVNIEKVPESYMENNHPDEYLVQVEHDFSGLRKHARRVISSNFLK 299
            EVL RHLG+  IDTVLVNIEKVP+ YM +N FDEYLVQVEHDF GL K   RVISSNFL+
Sbjct: 241 EVLHRHLGRPPIDTVLVNIEKVPQEYMNSNRFDEYLVQVEHDFVGLCKQVSRVISSNFLR 300

Query: 300 LEKGGAFHHGDFVVEELMNLV 320
            LE GGAFH GD +V+ELM ++
Sbjct: 301 LENGGAFHDGDLIVDELMRII 321
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5535> which encodes the amino acid sequence <SEQ ID 5536>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 251/320 (78%), Positives = 284/320 (88%)

Query: 1   MRKPKITVIGGGTGIPVILKSLRLEDVEITAVVTVADDGGSSGELRSVMQLTPPGDLRNV 60
            M+ PK+TVIGGGTGI +ILKSLR E V+ITAVVTVADDGGSSGELR+ MQL PPGDLRNV
Sbjct: 1   MKNPKMTVIGGGTGISILLKSLRNEAVDITAVVTVADDGGSSGELRNAMQLAPPGDLRNV 60

Query: 61  LVALSDMPKFYEQIPQYRFAEGDGDFAGHPLGNLIIAGVAEMQGSTYNAMQSLTQPFHTT 120
```

```
              L+A+SDMPKFYE++PQYRF E DG  AGHPLGNLIIAG++EMQGSTYNA+Q LT+FPH T
Sbjct: 61  LLAMSDMPKFYRRVPQYRFNKSDGALAGHPLGNLIIAGISEMQGSTYNAIQILTKFPHIT 120

Query: 121 GKIYPSSEHPLTLHAVFKDGHEVVGESQLADYKGMIDHVVVTNTYNEETPTASRKVVDAI 180
              GKIYPSSE  LTLHAVFKDGHEV GES IA Y GMIDHVYVTNTYN++ P ASRKVV+AI
Sbjct: 121 GKIYPSSEQALTLHAVFKDGHEVAGESSIAKYPGMIDHVYVTNTYNDQKPQASRKVVEAI 180

Query: 181 LESDMIVLGPGSLFTSILPNLVIPEIKQALLSTRAEVAYVCNIMTQRGETEHPTDADHVE 240
              LESDMIVLGPGSLFTSILPNLVIPEIK+AL +T+AEV Y+CNIMTQ GETE F+DADHV
Sbjct: 181 LESDMIVLGPGSLFTSILPNLVIPEIKEALRQTKAEVVYICNIMTQYGETEQFSDADHVA 240

Query: 241 VLKRHLGQDAIDTVLVNIRKVPESYMENNHPDEYLVQVEHDPSGLRKHARRVISSNFLKL 300
              VL +HLG+D IDTVLVN+ KVP++YM +N FDEYLVQV+HDF+GL + A+RVISS FL+L
Sbjct: 241 VLNQHLGRDLIDTVLVNVAKVPQAYMNSNKFDEYLVQVDHDFAGLCRAAKRVISSYFLRL 300

Query: 301 EKGGAFHHGDTPVVEELMNLV 320
              E GGAFH G+ VVEELMNLV
Sbjct: 301 ENGGAFHDGNLVVEELMNLV 320
```

SEQ ID 5534 (GBS269) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 12; MW 35kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 5; MW 60.5kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1781

A DNA sequence (GBSx1888) was identified in *S.agalactiae* <SEQ ID 5537> which encodes the amino acid sequence <SEQ ID 5538>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2479(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB96619 GB:AJ400630 hypothetical protein [Streptococcus pneumoniae bacteriophage MM1]
 Identities = 209/303 (68%), Positives = 260/303 (84%)

Query: 1   MSFTVKVKEELLGHKSENKMELSAIIKMSGSLGLANHGLNLSITTENAKIARHIYSMLEE 60
            MSFTV VKEE+LG     ++ ELSAIIKMSGS+GL+  GL LS+  TENAK+ARH+Y
Sbjct: 1   MSFTVAVKEEILGQHHLEKHELSAIIKMSGSIGLSTSGLTLSVVTENAKLARHLYESFLH 60

Query: 61  HYHLQPEIKYHQKTNLRKNKVYTVFIEEKVDVILADLKLADAFFGIETGIEHSILDNDEN 120
             Y ++ EI++HQ++NLRKNKVYTVF +EKV  +L+DL LAD+FFG+ETGI+ +IL ++E
Sbjct: 61  FYEIKSEIRHHQRSNLRKNKVYTVFTDEKVQDLLSDLHLADSFFGLETGIDAILSDESA 120

Query: 121 GRAYLRGAFLSTGTVREPDSGKYQLEIPSVYLDHAQDLANLMKKFMLDAKVIEHKHGAVT 180
            GRAYL GAFL+ G++R+P+SGKYQLEI SVYLDHAQ +A+L+++F+LDAKV+E K GAVT
Sbjct: 121 GRAYLCGAFLANGSIRDPESGKYQLEISSVYLDHAQGIASLLQQFLLDAKVLERKKGAVT 180

Query: 181 YLQKAEDIMDFLIVIDAMEARDAFERIKMIRETRNDINRANNVETANIARTITASMKTIN 240
            YLQ+AEDIMDFLIVI AM+ARD FE +K++RETRND+NRANN ETANIART++ASMKTIN
Sbjct: 181 YLQRAEDIMDFLIVIGAMQARDDFERVKILRETRNDLNRANNAETANIARTVSASMKTIN 240

Query: 241 NIIKIMDTIGPDALPSDLRQVAQVRVAHPDYSIQQLADSLETPLSKSGVNHRLRKINKIA 300
            NI KI D +G + LP DL++VAQ+R+ HPDYSIQQ+ADSL TPL+KSGVNHRLRKINKIA
Sbjct: 241 NISKIKDIMGLKNLPVDLQEVAQLRIQHPDYSIQQLADSLSTPLTKSGVNHRLRKINKIA 300
```

```
Query: 301 DEL 303
           DEL
Sbjct: 301 DEL 303
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5539> which encodes the amino acid sequence <SEQ ID 5540>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1698(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 222/303 (73%), Positives = 269/303 (88%)

Query: 1   MSFTVKVKEELLGHKSENWMELSAIIKMSGSLGLANHGLNLSITTENAKIARHIYSMLEE 60
           MSFT KVKEEL+    + +   EL+AIIK+SGSLGLA+  L+LSITTENAKIAR+IYS++E+
Sbjct: 1   MSFTTKVKEELIHLSTGDNNELAAIIKLSGSLGLAHQSLHLSITTENAKIARYIYSLIED 60

Query: 61  HYHLQPEIKYHQKTNLRKNRVYTVFIEEKVDVILADLKLADASFGIETGIESRSILDNDEN 120
            Y + PEI+YHQKTNLRKNRVYTV++E+ V+ ILADLKLAD+PFG+ETGIE  +L +D
Sbjct: 61  AYVIVPEIRYHQKTNLRKNRVYTVVEQGVETILADLKLADSPFGLETGIEPQVLSDDNA 120

Query: 121 GRAYLRGAFLSTGTVREPDSGKYQLEIFSVYLDHAQDLANLMKKFMLDAKVIEHKHGAVT 180
            GR+YL+GAFL+ G++R+P+SGKYQLEI+SVYLDHAQDLA LM+KFMLDAK IEHK GAVT
Sbjct: 121 GRSYLKGAFLAAGSIRDPRSGKYQLEIYSVYLDHAQDLAQLMQKFMLDAKTIEHKSGAVT 180

Query: 181 YLQKAEDIMDFLIVIDAMEARDAFERIMIRETRNDINRANNVETANIARTITASMKTIN 240
           YLQKAEDIMDFLI+I AM  ++ FE IK++RE RNDINRANN RTANIA+TI+ASMKTIN
Sbjct: 181 YLQKAEDIMDFLIIIGAMSCKEDFEAIKLLRSARNDINRANNAETANIAKTISASMKTIN 240

Query: 241 NIIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQQIADSLETPLSKSGVMHRLRKINKIA 300
           NIIKIMDTIG ++LP +L+QVAQ+RV HPDYSIQQ+AD+LE P++KSGVMHRLRKINKIA
Sbjct: 241 NIIKIMDTIGLESLPIELCQVAQLRVKHPDYSIQQVADALEFPITKSGVMHRLRKINKIA 300

Query: 301 DEL 303
           D+L
Sbjct: 301 DDL 303
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1782

A DNA sequence (GBSx1889) was identified in *S.agalactiae* <SEQ ID 5541> which encodes the amino acid sequence <SEQ ID 5542>. This protein is predicted to be dipeptidase. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3544(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA86210 GB:Z38063 dipeptidase [Lactobacillus helveticus]
 Identities = 218/473 (46%), Positives = 310/473 (65%), Gaps = 14/473 (2%)
```

```
Query: 3    CTTILVGKKASYDGSTMIARTRDSVNGDPTPKKLKVMTSKDQPRHYKSVLSNPEVD---L 59
            CTTILVGKKAS DGSTMIAR+ED        P+  KV+  +DQP+HY SV+S  ++D   L
Sbjct: 6    CTTILVGKKASIDGSTMIARSEDG-GRVIIPBGPKVVNPEDQPKHYTSVISKQKIDDEDL 64

Query: 60   PDNDLPYTSVPDALGKDGIWGEAGINSKNVAMSATETTTTNSRVLGADPLVSD---GIGE 116
             + PL YTS PD  GK+GIWG AGIN+ NVAM+ATETTTTNSR+ G DP++     G+GE
Sbjct: 65   AETPLRYTSAPDVSGKNGIWGAAGINADNVAMTATETTTTNSRIQGVDPILDPSBGGLGE 124

Query: 117  EDILTLVLPYIQSARBGVERLGAILEKYGTYESNGIAFSDTREIWWLETIGGHHWIARRV 176
             ED +TL LPY+ SA +GV+R+G ++EKYGTYS NG+AFSD + IW+LETIGGHHWIARR+
Sbjct: 125  EDFVTLTLPYLHSAPDGVKKVGYLVEKYGTYEMNGMAFSDKDNIWYLETIGGHHWIARRI 184

Query: 177  PDDVYVTNPNQLGIDHFEFNNCDDYMCSSDLKSPIBCYHLDLTYSNEHFNPRYAPGSQRD 236
             PDD YV  PN+L ID F+F++ +++   +SDLK+ I++YHL+     E +N R+ PGS
Sbjct: 185  PDDAYVIAPNRLNIDTFDFDDSENFAAASDLKDLIDEYHLN--PDREGYNMRHIPGSSTI 242

Query: 237  KDRHYNTPRSWAMQRFLNPEIEQDPRSLFIPWCQKPYRKITVEDIKYVLSDHYQDSVYDP 296
             KD HYN PR+W +  + +P+    P      P+  +  R I++EDIX+  S HYQD+ YD
Sbjct: 243  KDAHYNNPRAWYIHNYFDPDFGGTPADQDQPFICRANRLISIEDIKWABSSHYQDTPYDA 302

Query: 297  YGPRGDAVSRRAFRSVGINRTSQTSILQLRNKSLETTGVQWLSYGSMPFATMVPLFTGV 356
             YG +G     ++ FR +GINR  +T ILQ+R +   E  GVQWL++G   F +M+P +T V
Sbjct: 303  YGDQSTPEQKKTFRPIGINRNFETHILQIRNDVFAELAGVQWLAPGPNTPNSMLPFTNV 362

Query: 357  ETVPNYFSNTTKDASTDNFYWTNRLIAALADPHFYQHRADIKSYIERTWAQGHAHINGVD 416
              T P  +  T K  + +  +W N+L A L D ++  +     +++ +++++AQ H  +  D
Sbjct: 363  TTTPEANQTTPK-FNLNKIFWLNKLTAQLGDTNYRVYGKLEDAFBQKSLAQCHKIQHETD 421

Query: 417  REVAENKEIDPQQK----NQEMSDYIQKESQELLNRILFDASNLMTNRFSMGD 465
             +EV      + Q K    NQ+MSD +   + ELL +++ +   LMT ++ + D
Sbjct: 422  KEVKNLSGIKELQDKLIAANQKMSDIVYNNTVRLLGQMVDKGHGLMTLKYDLLD 474
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5543> which encodes the amino acid sequence <SEQ ID 5544>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0514(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 345/464 (74%), Positives = 407/464 (87%)

Query: 2    ACTTILVGKKASYDGSTMIARTRDSVNGDFTPKKLKVMTSKDQPRHYKSVLSNPEVDLPD 61
            +CTTILVGKKASYDGSTM+ARTEDS NGDFTPKK+ V+  +DQPRHY+SV S+FE+DLPD
Sbjct: 9    SCTTILVGKKASYDGSTMVARTEDSQNGDFTPKKMIVVKPEDQPRHYRSVQSSFEMDLPD 68

Query: 62   NPLPYTSVPDALGKDGIWGEAGINSKNVAMSATETTTTNSRVLGADPLVSDGIGEEDILT 121
            NP+ YTSVPDALGKDGIW EAG+N  NVAMSATETTTTNSRVLGADPLV+ GIGEED++T
Sbjct: 69   NPMTYTSVPDALGKDGIWAEAGVNEANVAMSATETTTTNSRVLGADPLVASGIGEEDMVT 128

Query: 122  LVLPYIQSARBGVERLGAILEKYGTYESNGIAFSDTREIWWLETIGGHHWIARRVPDDVY 181
            LVLPYI+SAREGV RLGAILE YGTYESNG+AFSD  +IWWLETIGGHHWIARRVPDD Y
Sbjct: 129  LVLPYIRSAREGVLRLGAILEDYGTYESNGVAFSDRHDIWWLETIGGHHWIARRVPDDAY 188

Query: 182  VTNPNQLGIDHFEFNNCDDYMCSSDLKEFIEQYHLDLTYSNEHFNPRYAPGSQRDKDRHY 241
            VTNPNQ GIDHFEFNN +DY+CS+DLK+FI+  YHLDLTYS+EHFNPRYAPGSQRDKDR Y
Sbjct: 189  VTNPNQFGIDHFEFNNPEDYLCSADLKDFIDTYHLDLTYSHEHFNPRYAPGSQRDKDRQY 248

Query: 242  NTPRSWAMQRFLNPEIEQDPRSLFIPWCQKPYRKITVEDIKYVLSDHYQDSVYDPYGPEG 301
            NTPR+W MQ+FLNPEI QDPRS  + WCQKPYRKITVED+KYVLS HYQD+ YDPYG EG
Sbjct: 249  NTPRAWIMQKFLNPEIVQDPRSFALAWCQKPYRKITVEDVKYVLSSHYQDTGYDPYGSEG 308
```

```
Query: 302 DAVSRRAFRSVGINRTSQTSILQLRPNKSLETTGVQWLSYGSMPFATMVPLFTQVETVPN 361
            VS++ FR +GINRTSQT+XL +RPNK  E   +QW++YGSMPF TMVP FTQV+T+P+
Sbjct: 309 TPVSKKVFRPIGINRTSQTAILHIRPNKDQEIAAIQWMAYGSMPFTMVPFFTQVKTIPD 368

Query: 362 YFSNITKDASTDNFYWTNRLIAALADPHFYQHEADIESYIERTMAQGHAHINGVDREVAE 421
            YF+NT ++  TDNFYWTNRLIAALADPH+  HE D+++Y+E TMA+GHA ++ V+ ++
Sbjct: 369 YFANTYENVFTDNFYWTNRLIAALADPHYNIHETDLDNYLESTMAKGHAMLHAVEVQLLA 428

Query: 422 NKEIDPQQKNQRMSDYIQKESQELLNRILFDASNLMTNRFSMGD 465
             + +D +++NQ+MSDY+Q E+Q LLN+ILFDASNLMTNRFS+ D
Sbjct: 429 GETVDLEEENQKMSDYVQGETQTLLNKILFDASNLMTNRFSLSD 472
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1783

A DNA sequence (GBSx1890) was identified in *S.agalactiae* <SEQ ID 5545> which encodes the amino acid sequence <SEQ ID 5546>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA96185 GB:Z71552 AdcA protein [Streptococcus pneumoniae]
 Identities = 257/429 (59%), Positives = 312/429 (71%), Gaps = 7/429 (1%)

Query: 1   MRKKFLLLMSPVAMFAAWQLVQVKQVWADSKLKVVTFPYVYEFTKNVVGDKADVSMLIK 60
            M+K  LLL S  A+F    +    Q  AD KL +VTFPYVYEFTK V GD A+V +LI
Sbjct: 1   MKKISLLLASLCALPL---VACSNQKQADGKLNIVTFPYVYEFTKQVAGDTANVELLIG 57

Query: 61  AGTEPHDFEPSTKNIAAIQDSNAFVYNDDNMETWAPKVAKSVKSKKVTTIKGTGDMLLTK 120
            AGTEPH++EPS K +A IQD++ FVY ++NMETW PK+  ++  KKV TIK TGDMLL
Sbjct: 58  AGTEPHEYEPSAKAVAKIQDADTFVYENENMETWVPKLLDTLDKKKKVKTIKATGDMLLLP 117

Query: 121 GVEEBGEEHEGHGHEGHHHELDPHVWLSPERAISVVENIRNKFVKAYPKDAASFNKNADA 180
            G EEE  +H+ HG EGHHHE DPHVWLSP RAI  +VE+IR+     YP     +F KNA A
Sbjct: 118 GGEEEEGDHD-HGEEGHHHEPDPHVWLSPVRAIKLVEHIRDTLSADYPDKKETFEKNAAA 176

Query: 181 YIAKLKELDKEYKNGLSNAKQKGFVTQHAARGYMALDYGLNQVPIAGLTPDAEPSSKRLG 240
            YI KL+ LDK Y  GLS AK+KGFVTQHAAF Y+ALDYGL QV I+GL+PDAEPS+ RL
Sbjct: 177 YIEKLQSLDKAYABGLSQAKEKSFVTQHAAFNYLALDYGLKQVAISGLSPDAEPSAARLA 236

Query: 241 EIAKYIKKYNINYIYPEENASNKVAKTLADEVGVKTAVLSPLSGLSKKEMAAGEDYFSVM 300
            EL +Y+KK  I YIYPEENAS  +A TL+ E GVKT VL+PLE L++++  AGE+Y SVM
Sbjct: 237 ELETYVKKNKIAYIYPEENASQALANTLSKRAGVKTDVLNPLESLITEKTYKAGENYISVM 296

Query: 301 RRNLKVLKKTIDVAGKEVAPEE-DKTKTVETGYFKTKDVKDRKLTDYSGNWQSVYPLLQD 359
            +NLK LK+TID  G  + PE+ + TKTV+ GYF+   VKDR L+DY+GNWQSVYP L+D
Sbjct: 297 KENLKALKQTIDQHGPAIEPEKAEDTKTVQNGYPETAAVKDRTLSDYAGNWQSVYPFLED 356

Query: 360 GTLDPVWDYKAKSKKDMTAAEYKKYYTAGYKTDVESIKIDGKKHQMTFVRNGKSQTTTYK 419
            GT D V+DYKAK    MT AEYK YYT GV+TDV  I I    + M FV+ G+S+ +TYK
Sbjct: 357 GTFDQVFDYKAKLTGKMTQAEYKAYYTKGVQTDVTKINI--TDNTMEFVQGQGSKKYTYK 414

Query: 420 YAGYKILTY 428
            Y G KILTY
Sbjct: 415 YVGKKILTY 423
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5547> which encodes the amino acid sequence <SEQ ID 5548>. Analysis of this protein sequence reveals the following:

```
        Possible site: 17

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA96185 GB:Z71552 AdcA protein [Streptococcus pneumoniae]
 Identities = 259/438 (59%), Positives = 326/438 (74%), Gaps = 16/438 (3%)

Query: 1   MKKKILLMMSLISVFPAWQLTQAKQVLADGKVKVVTTFYPVYVEFTKGVIGNDGDVFNLMK 60
           MKK  LL+ SL ++F    + + Q  A+GK+ +VTFFYPVYVEFTK V G+  +V +L+
Sbjct: 1   MKKISLLLASLCALFL---VACSNQKQADGKLNIVTFFYPVYVEFTKQVAGDTANVELLIG 57

Query: 61  AGTEPHDFEPSTKDIKKIQDADAFVYMDDNMETWVSDVKKSLTSKKVTIVKGTGNMLLVA 120
           AGTEPH++EPS K + KIQDAD FVY ++NMETWV  +  +L  KKV  +K TG+MLL+
Sbjct: 58  AGTEPHEYEPSAKAVAKIQDADTFVYENENMETWVPKLLDTLDKKKVKTIKATGDMLLLP 117

Query: 121 GAGHDHPHEDADKKHEHNHKHSEEGHNHAPDPHVWLSPYRSITVVENIRDSLSKAYPEKAE 180
           G     E+ +  H+H    EEGH+H FDPHVWLSP R+I +VE+IRD+LS  YP+K E
Sbjct: 118 GG------EEEEGDHDHG---EEGHHHEPDPHVWLSPVRAIKLVEHIRDTLSADYPDKKE 168

Query: 181 NFKANAATYIEKLKRLDKDYTAALSDAKQKSFVTQHAAFGYMALDYGLNQISINGVTPDA 240
            F+ NAA YIEKL+ LDK Y   LS AK+KSFVTQHAAF Y+ALDYGL Q++I+G++PDA
Sbjct: 169 TFEKNAAAYIEKLQSLDKAYAEGLSQAKEKSFVTQHAAFNYLALDYGLKQVAISGLSPDA 228

Query: 241 EPSAKRIATLSKYVKKYGIKYIYFEENASSKVAKTLAKEAGVKAAVLSPLEGLTEKEMKA 300
           EPSA R+A L+ YVKK  I YIYFEENAS  +A TL+KEAGVK  VL+PLE LTS++ KA
Sbjct: 229 EPSAARLAELTEYVKKNKIAYIYFEENASQALANTLSKEAGVKTDVLNPLESLTSEDTKA 288

Query: 301 GQDVFTVMRKNLETLRLTTDVAGKEILPEK-DTTKTVVMGYPKDKEVKDRQLSDMSGSMQ 359
           G++Y +VM KNL+ L+ TTD  G  I PEK + TKTV NGYF+D  VKDR L+D++G+WQ
Sbjct: 289 GENYISVMEKNLKALKQTTDQEGPAIEPEKAEDTKTVQNGYFRDAAVKDRTLSDYAGNWQ 348

Query: 360 SVYPYLQDGTLDQVWDVKAKKSKGKMTAAEYKDYTTPGYKTDVEQIKINGKKKTMTFVRN 419
           SVYP+L+DGT DQV+DYKAK + GKMT AEYK YYT GY+TDV   KIN    TM FV+
Sbjct: 349 SVYPFLEDGTPDQVFDYKAKLT-GKMTQAEYKAYYTKGYQTDV--TKINITDNTMEFVQG 405

Query: 420 GEKKTFTTYYAGKEILTY 437
           G+ K +TY Y GK+ILTY
Sbjct: 406 GQSKKYTKYVGKKILTY 423
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 353/515 (68%), Positives = 422/515 (81%), Gaps = 9/515 (1%)

Query: 1   MRKKFLLLMSFVAMFAAWQLVQVKQVWADSKLKVVTTFYPVYVEFTKNVVGDKADVSNLIK 60
           M+KK LL+MS +++F AWQL Q KQV A+ K+KVVTTFYPVYVEFTK V+G+  DV ML+K
Sbjct: 1   MKKKILLMMSLISVFPAWQLTQAKQVLADGKVKVVTTFYPVYVEFTKGVIGNDGDVFNLMK 60

Query: 61  AGTEPHDFEPSTKNIAAIQDSNAFVYMDDNMETWAPKVAKSVKSKKVTTIKGTSDMLLTK 120
           AGTEPHDFEPSTK+I  IQD++AFVYMDDNMETW   V KS+ SKKVT +KGTG+MLL
Sbjct: 61  AGTEPHDFEPSTKDIKKIQDADAFVYMDDNMETWVSDVKKSLTSKKVTIVKGTGNMLLVA 120

Query: 121 GV--------EEEGRHHEGHGHGHEHHHELDPHVWLSPRRAISVVENIRNKFVKAYPKDAA 172
           G          ++  EH  H  EGH+H  DPHVWLSP R+I+VVENIR+   KAYP+ A
Sbjct: 121 GAGHDHPHEDADKKHEHNHKHSEEGHNHAPDPHVWLSPYRSITVVENIRDSLSKAYPEKAE 180

Query: 173 SFNKNADAYIAKLKELDKEYEKGLSNAKQKSFVTQHAAFGYMALDYGLNQVPIAGLTPDA 232
           +F  NA  YI KLKELDK+Y   LS+AKQKSFVTQHAAFGYMALDYGLNQ+ I G+TPDA
```

```
Sbjct: 181 NFKANAATYIEKLKELDKDYTAALSDAKQKSFVTQHAAFGYMADLYGLNQISNGVTPDA 240

Query: 233 EPSSKRLGELAKYIKKYNINYIYFSENASNKVAKTLADEVGVKTAVLSPLEGLSKKEMAA 292
           EPS+KR+  L+KY+KKY I YIYFSENAS+KVAKTLA E GVK AVLSPLEGL++KEM A
Sbjct: 241 EPSAKRIATLSKYVKKYGIKYIYFSENASSKVAKTLAKEAGVKAAVLSPLEGLTEKEMKA 300

Query: 293 GEDYPSVMRRNLKVLKKTTDVAGKEVAPEEDKTKTVSTGYPKTKDVKDRKLTDYSGNWQS 352
           G+DYP+VMR+NL+ L+  TTDVAGKE+ PE+D TKTV  GYPK K+VKDR+L+D+SG+WQS
Sbjct: 301 GQDYPTVMRKNLETLRELTTDVAGKELPEKDTTKTVNNGYPEDKEVKDRQLSDWSGSWQS 360

Query: 353 VYPLLQDGTLDPVWDYKA-KSKKDWTAAEYKKYTAGYKHDVESIKIDGKKHQMTFVRNG 411
           VYP LQDGTLD VWDYKA KSK  MTAAEYK YTT GYKTDVE IKI+GKK  MTFVRNG
Sbjct: 361 VYPVLQDGTLDQVWDYKAKKSKGKMTAAEYKDYTTGYKHDVEQIKINGKKKTMTFVRNG 420

Query: 412 KSQTPTYKYAGYKILTYKKGNRGVRYLFEAKEKDAGQFKYIQFSDHGIKPNKAEHFHIFW 471
           + +TFTY YAG +ILTY KGNRGVR++FEAKE DAG+FKY+QFSDH I P KA+HPH++W
Sbjct: 421 EKKTFTYTYAGKEILTYPKGNRGVRFMFEAKEADAGEFKYVQFSDHAIAPEKAKHFHLYW 480

Query: 472 GSESQEKLFEEMENWPTYFPAKMSGR6VAQDLMSH 506
           G +SQEKL +E+E+WPTY+ + +SGRE+AQ++ +H
Sbjct: 481 GDDSQEKLHKELEHWPTYYGSDLSGRELAQEINAH 515
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8899> and protein <SEQ ID 8900> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length of UR:    19
     Peak Value of UR:    2.79
     Net Charge of CR: 3
McG: Discrim Score:       9.08
GvH: Signal Score (-7.5): 2.59
     Possible site: 15
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 16
ALOM program   count: 0 value:   7.69 threshold:   0.0
   PERIPHERAL  Likelihood =  7.69     264
 modified ALOM score:  -2.04

*** Reasoning Step: 3

Rule gpo1

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
3758895|emb|CAA96185.1||Z71552 AdcA protein {Streptococcus pneumoniae} >PIR|T46756|T46756
Zn-binding lipoprotein
          adcA [imported] - Streptococcus pneumoniae (fragment)

 Score =  508 bits (1294), Expect = e-143
 Identities = 257/429 (59%), Positives = 312/429 (71%), Gaps = 7/429 (1%)

Query: 1   MRKKFLLLMSFVAMFAAWQLVQVKQVWADSKLKVVTFFYPVYRFTKNVVGDKADVSMLIK 60
           M+K  LLL S  A+F    +     Q  AD KL +VT FYPVYEFTK V GD A+V +LI
Sbjct: 1   MKKISLLLASLCALFL---VACSNQKQADGKLNIVTTFYPVYKFTQVAGDTANVELLIG 57

Query: 61  AGTEPHDFEPSTKNIAAIQDSNAFVYMDDNMETWAPKVAKSVKSKKVTTIKGTGDMLLTK 120
           AGTEPH++EPS K +A IQD++ FVY ++NMETW PK+  ++  KKV TIK TGDMLL
```

```
Sbjct: 58  AGTEPHEYEPSAKAVAKIQDADTFVYENENMSTWVPKLLLTLDKKKVKTIKATGDMLLLP 117

Query: 121 GVEEBGEEHEGHGHEGHHHELDPHVWLSPERAISVVENIRNKFVKAYPKDAASFNKNADA 180
           G EEE  +H+ HG EGHHHR DPHVWLSP RAI +VB+IR+     YP    +F KNA A
Sbjct: 118 GGRERKDHD-HGEEGHHHERDPHVWLSPVRAIKLVEHIRDITLSADYPDKKETFRKNAAA 176

Query: 181 YIAKLKELDKEYKNGLSNAKQKSFVTQHAAFGYMALDYGLNQVPIAGLTPDAEPSSKRLG 240
           YI KL+ LDK Y  GLS AK+KSFVTQHAAF Y+ALDYGL QV I+GL+PDAEPS+ RL
Sbjct: 177 YIEKLQSLDKAYAEGLSQAKEKSFVTQHAAFNYLALDYGLKQVAISGLSPDAEPSAARLA 236

Query: 241 ELAKYIKKYNINYIYFEENASNKVAKTLADEVGVKTAVLSPLEGLSKKEMAAGEDYFSVM 300
           EL +Y+KK  I YIYFEENAS  +A TL+ E GVKT VL+PLE L++++  AGE+Y SVM
Sbjct: 237 ELTEYVKKNKIAYIYFEENASQALANTLSKEAGVKTDVLNPLESLTEEDTKAGENYISVM 296

Query: 301 RRNLKVLKETTDVAGKEVAPER-DKTKTVRTGYFKTKDVKDRKLTDYSGNWQSVYPILQD 359
            +NLK LK+TTD  G  + PE+ + TKIV+ GYF+    VKDR L+DY+GNWQSVYP L+D
Sbjct: 297 EKNLKALKQTTDQSGPAIEPEKAEDTKTVQNGYFEDAAVKDRTLSDYAGNWQSVYPFLED 356

Query: 360 GTLDPVWDYKAXSKKDWTAAEYKKYYTAGYKTDVESIKIDGKKHQMTFVRNGKSQTFTYK 419
           GT D V+DYKAK    MT AEYK YYT GY+TDV  I I    + M FV+ G+S+ +TYK
Sbjct: 357 GTFDQVFDYKAKLTGKMTQAEYKAYYTKGYQTDVTKINI--TDNTMEFVQSGQSKKYTYK 414

Query: 420 YAGYKILTY 428
           Y G KILTY
Sbjct: 415 YVGKKILTY 423
```

SEQ ID 8900 (GBS325) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 65 (lane 3; MW 58kDa).

The GBS325-His fusion product was purified (Figure 210, lane 7) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 257A) and FACS (Figure 257B). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Example 1784

A DNA sequence (GBSx1891) was identified in *S.agalactiae* <SEQ ID 5549> which encodes the amino acid sequence <SEQ ID 5550>. This protein is predicted to be ribosomal protein L31 (rl31). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1948(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9637> which encodes amino acid sequence <SEQ ID 9638> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF80389 GB:AF160251 ribosomal protein L31 [Listeria innocua]
 Identities = 61/81 (75%), Positives = 71/81 (87%), Gaps = 1/81 (1%)

Query: 9   MKKDIHPDYRPVVFLDTTTGYKFLSGSTKSTKEYVEPE-GETYYLLRVEISSDSHPFYTG 67
           MK  IHP+YRPVVF+DT+T +KFLSGSTKS+ ET+++E G  YPL+RVEISSDSHPFYTG
Sbjct: 1   MKTGIHPEYRPVVFVDTSTDFKFLSGSTKSSSETIKWEKGNEYPLLRVEISSDSHPFYTG 60

Query: 68  RQKFTQADGRVDRFNKKYGLK 88
           +QK   ADGRVDRFNKKYGLK
Sbjct: 61  KQKHACADGRVDRFNKKYGLK 81
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5551> which encodes the amino acid sequence <SEQ ID 5552>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1910(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 81/86 (94%), Positives = 86/86 (99%)

Query: 9  MKKDIHPDYRPVVFLDTTGYKFLSGSTKSTKETVEFEGETYPLIRVEISSDSHPFYTGR 68
          M+KDIHPDYRPVVFLDTTGYQFLSGSTKASKETVEFEGETYPLIRVEISSDSHPFYTGR 60

Query: 69 QKFTQADGRVDRFNKXYGLKDANAAQ 94
          QKFTQADGRVDRFNKXYGLKDANAA+
Sbjct: 61 QKFTQADGRVDRFNKXYGLKDANAAK 86
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1785

A DNA sequence (GBSx1892) was identified in *S.agalactiae* <SEQ ID 5553> which encodes the amino acid sequence <SEQ ID 5554>. This protein is predicted to be aspartate aminotransferase (aspC). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1740(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9421> which encodes amino acid sequence <SEQ ID 9422> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC21948 GB:U32714 aminotransferase [Haemophilus influenzae Rd]
 Identities = 200/323 (61%), Positives = 264/323 (80%), Gaps = 1/323 (0%)

Query: 1   MQYYQLQNI-HVDMDDIYIVNGVSEGISMSMQALLDNDDEVLVPMPDYPLWTACVSLAGG 59
           +QYYQ + I    ++D+YI NGVSE I+M+MQALL++ DEVLVPMPDYPLWTA V+L+GG
Sbjct: 82  VQYYQSKGILGATVNDVYIGNGVSELITMAMQALLNDGDEVLVPMPDYPLWTAAVTLSGG 141

Query: 60  NAVHYICDSEANWYPDIDDIKSKITSRTKAIVLINPNNPTGAVYPRSILQEIVDIARQND 119
            AVHY+CDE+ANW+P IDDIK+K+ +KTKAIV+INPNNPTGAVY +E+LQEIV+IARQN+
Sbjct: 142 KAVHYLCDEDANWFPTIDDIKAKVNAKTKAIVIINPNNPTGAVYSKSLLQEIVEIARQNN 201

Query: 120 LIIFSDEVYDRLVMDGMEHIPIASIAEDIFTVTLSGLSKSHRICGFRVGWMVLSGPRQHV 179
           LIIF+DE+YD++  DG  H  IA++A D+ TVTL+GLSK++R+ GFR GWM+L+GP+ +
Sbjct: 202 LIIFADEIYDKILYDGAVHHHIAALAPDLLTVTLNGLSKAYRVAGFRQGWMILNGPKHNA 261

Query: 180 KGYIEGLNMLANMRLCSNVLAQQVIQTSLGGQQSIDSMLLPGRIYEQRNYIHKAINEIP 239
           KGYIEGL+MLA+MRLC+NV  Q  IQT+LGG QSI+  +LPGGR+ EQRN  +  I +IP
Sbjct: 262 KGYIEGLDMLASMRLCANVPMQHAIQTALGGYQSINEFILPGGRLLEQRNKAYDLITQIP 321
```

```
Query: 240 GLSAVKPNAGLYLPPKIIYTDMYRIDNDKRPVLNFLKQRKVLLTHGRGFNMNTADHFRIVY 299
           G++ VKP    +Y+PPKID    + I +DE+ VL+ L+QRKVLL HG+GFN ++ DHFRIV
Sbjct: 322 GITCVKPMGAMYMPPKIDVKKFNIHSDEKMVLDLLRQRKVLLVHGKGFNWHSPDHFRIVT 381

Query: 300 LPRVDELTELQEKMARFLSQYKR 322
           LP V++L E   K+ARFLS Y++
Sbjct: 382 LPYVNQLEEAITKLARFLSDYRQ 404
```

There is also homology to SEQ ID 3662.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1786

A DNA sequence (GBSx1893) was identified in *S.agalactiae* <SEQ ID 5555> which encodes the amino acid sequence <SEQ ID 5556>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.02   Transmembrane  164 - 180 ( 163 - 181)

----- Final Results -----
                bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10099> which encodes amino acid sequence <SEQ ID 10100> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06181 GB:AP001515 transcriptional pleiotropic repressor
           [Bacillus halodurans]
 Identities = 129/257 (50%), Positives = 181/257 (70%), Gaps = 3/257 (1%)

Query: 23  NLLEKTKKITSILQRSVDSLDAELPYNTMAQLADIIDCNACIINGGGNLLGYAMKYKTN 82
           +LL + RKI  +LQ+S      + +  MA  L D+I  N  +++  G LLG+A+K +
Sbjct: 2   SLLSRMRKINDMLQKSGVQ---HVNFREMAETLRDVISANIFVVSRRGKLLGFAIKQEIE 58

Query: 83  TDRVEEFFETKQPFDYYVKSASKRVYDTEANLSVDNDLSIFPVETKENPQDGITTIAPIYG 142
            +R+++  E +QFP+ Y      +V +T ANL ++++ + FPVE KE F+ G+TTI PI G
Sbjct: 59  NERMKMLEDRQPFEEYTTGLFKVETSANLDINSEFTAFPVENKELFKTGLTTIVPISG 118

Query: 143 GGMRLGTFIIWRNDKEFSDDDLILVEIASTVVGIQLLNLQTENLENIRKQTAVTMAINT 202
           GG RLGT I+ R +  F+DDDLIL E  +TVVG+++L+ +T+ +EE  R +  V MAI++
Sbjct: 119 GGQRLGTLILARINDSFNDDDLILAEYGATVVGMEILHEKTQEIEERARSKAVVQMAISS 178

Query: 203 LGVSEMKAVAAILGKLDGLEGKLTASVIADRIGITRSVIVNALRKLESAGIIESRSLGMK 262
           LGVSE+KAV  I  ELDG EG L AS IADR+GITRSVIVNALRKLESAG+IESRSLGMK
Sbjct: 179 LGVSELEAVEHIFEELDGKEGLLVASKIADRVGITRSVIVNALRKLESAGVIESRSLGMK 238

Query: 263 GTYLKVINEGIFDKLKK 279
           GTY+KV+N+    +L++
Sbjct: 239 GTYIKVLNDKFLVELEK 255
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5557> which encodes the amino acid sequence <SEQ ID 5558>. Analysis of this protein sequence reveals the following:

```
      Possible site: 38
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.64   Transmembrane  144 - 160 ( 143 - 161)

----- Final Results -----
```

```
bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB13490 GB:Z99112 transcriptional regulator [Bacillus subtilis]
 Identities = 131/255 (51%), Positives = 179/255 (69%), Gaps = 3/255 (1%)

Query: 4   LLEKTRKITSILQRSVDSLETELPYNTMASRLADIIDCNACIINGGGTLLGYAMKYKTNT 63
            LL+KTR I S+LQ +       + +  MA  L D+ID N  +++   G LLGY++  +
Sbjct: 3   LLQKTRIINSMLQAAAGK---PVNFKEMAETLRDVIDSNIPVVSRRGKLLGYSINQQIEN 59

Query: 64  DRVEEPFEAKQFPDTYVKAASRVYDTEANLSVENELTIFPVESKDTYPGGLTTIAPIYGG 123
            DR+++  E +QFP+ Y K    V +T +NL + +E T FPVE++D +  GLTTI PI GG
Sbjct: 60  DRMKKMLEDRQFPEEYTKNLFNVPETSSNLDINSEYTAFPVENRDLPQAGLTTIVPIIGG 119

Query: 124 GMRLGSLIIWKNDNEPSDDDLILVEISSTVVGIQLLNLQTENLEDTIRKQTAVNMAINTL 183
            G RLG+LI+ R  ++F+DDDLIL E  +TVVG++LL  + E +E+  R +  V MAI++L
Sbjct: 120 GERLGTLILSRLQDQFNDDDLILAEYGATVVGMELLREKAEEIEEEARSKAVVQMAISSL 179

Query: 184 SYSEMKAVAAILGELDGNEGRLTASVIADRIGITRSVIVNALRKLESAGIIESRSLGMKG 243
            SYSE++A+  I  ELDGNEG L AS IADR+GITRSVIVNALRKLESAG+IESRSLGMKG
Sbjct: 180 SYSELEAIEHIFEELDGNEGLLVASKIADRVGITRSVIVNALRKLESAGVIESRSLGMKG 239

Query: 244 TYLKVINEGIFAKLK 258
            TY+KV+N     +L+
Sbjct: 240 TYIKVLNNKFLIELE 254
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 232/260 (89%), Positives = 247/260 (94%)

Query: 21  MPNLLEKTRKITSILQRSVDSLDAELPYNTMAAQLADIIDCNACIINGGGNLLGYAMKYK 80
            MPNLLEKTRKITSILQRSVDSL+ ELPYNTMA++LADIIDCNACIINGGG LLGYAMKYK
Sbjct: 1   MPNLLEKTRKITSILQRSVDSLETELPYNTMASRLADIIDCNACIINGGGTLLGYAMKYK 60

Query: 81  TNTDRVEEPFETKQFPDYVVKASRVYDTEANLSVDNDLSIPVVETKENPQDGITTIAPI 140
            TNTDRVEEPFE KQFPD YVK+ASRVYDTEANLSV+N+L+IPVVE+K+ +  G+TTIAPI
Sbjct: 61  TNTDRVEEPFEAKQFPDTYVKAASRVYDTEANLSVSNELTIPVVESKDTYPGGLTTIAPI 120

Query: 141 YGGGMRLGTFIIWRNDKEFSDDDLILVEIASTVVGIQLLNLQTENLEENIRKQTAVTMAI 200
            YGGGMRLG+ IIWRND EFSDDDLILVEI+STVVGIQLLNLQTENLE+ IRKQTAV MAI
Sbjct: 121 YGGGMRLGSLIIWRNDNEFSDDDLILVEISSTVVGIQLLNLQTENLEDTIRKQTAVNMAI 180

Query: 201 NTLSYSEMKAVAAILGELDGLEGRLTASVIADRIGITRSVIVNALRKLESAGIIESRSLG 260
            NTLSYSEMKAVAAILGELDG EGRLTASVIADRIGITRSVIVNALRKLESAGIIESRSLG
Sbjct: 181 NTLSYSEMKAVAAILGELDGNEGRLTASVIADRIGITRSVIVNALRKLESAGIIESRSLG 240

Query: 261 MKGTYLKVINEGIFDKLKEY 280
            MKGTYLKVINEGIF KLKE+
Sbjct: 241 MKGTYLKVINEGIFAKLKEF 260
```

A related GBS gene <SEQ ID 8901> and protein <SEQ ID 8902> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:       -6.84
GvH: Signal Score (-7.5): -5.37
     Possible site: 13
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -2.02 threshold:  0.0
   INTEGRAL    Likelihood =  -2.02   Transmembrane  114 - 130 ( 113 - 131)
   PERIPHERAL  Likelihood =  3.61     179
 modified ALOM score:   0.90

*** Reasoning Step: 3
```

```
----- Final Results -----
                bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02556(223 - 987 of 1293)
EGAD|13275|BS1617(4 - 255 of 259) codY protein {Bacillus subtilis} OMNI|NT01BS1895 codY
protein (vegetative protein 286b) (veg286b) GP|535351|gb|AAB03372.1||U13634 CodY {Bacillus
subtilis} GP|2633989|emb|CAB13490.1||Z99112 transcriptional regulator {Bacillus subtilis}
PIR|S61496|S61496 transcription pleiotropic repressor codY - Bacillus subtilis
%Match = 29.1
%Identity = 50.6  %Similarity = 71.5
Matches = 128  Mismatches = 71  Conservative Sub.s = 53

177       207       237       267       297       327       357       387
DCKS*NALI*L*RKTYKG*RKCRIYLEKTRKITSILQRSVDSLDAELPYNTMAAQLADIIDCNACIINGGGNLLGYAMKY
                              |:||| | |:|| :       : :  ||  | |:|| |  :::  | ||||::
                            MALLQKTRIINSMLQAAAGK---PVNFKEMARTLRDVIDSNIFVVSRRGKLLGYSINQ
                                      10        20        30        40        50

417       447       477       507       537       567       597       627
KTNYDRVEEPETKQFDDYTVKSASRVVDTEANLSVDNDLSIFPVETKENRQDQITTIAPIYGGKMRLGTFIIWRNDKEF
:   ||::: :| :|||: | |:   | :|  :|! :::: : ||||  :: || |:||| || ||| |||:|:|:  |   :|
QIENDNMKKMGLERQFPEEYTNLVNVPTSSNLDINSEYTAPFVENRDLAQGLTTTVPIIGGGERLGTLILSRLQDQP
          70        80        90       100       110       120       130

657       687       717       747       777       807       837       867
SDDDLILVEIASTVVQIQLLNLQTENLRENIRKQTAVTMAINTLSYSEMKAVAAILGELDGLEGRLTASVIADRIGITRS
:|||||| |  :||||:::|  : | :||  | :  | |||::|||||::|:  |: |||  || | || ||||:|||||
NDDDLILAEYGATVVGMEILREKAEEIEEARSKAVVQMAISSLSYSELEAIEHIFEELDGNEGLLVASKIADRVGITRS
          150       160       170       180       190       200       210

897       927       957       987      1017      1047      1077      1107
VIVNALRKLESAGIIESRSLGMKGTYLKVINEGIPDKLKEYN*S*HGTGSSFQFLFWNQEEIRRKMTXXN*LXXLFS*RL
|||||||||||||:||||||||||||||||:||:|  : :|:
VIVNALRKLESAGVIRSRSLGMKGTYIKVLNNKFLIELENLKGH
          230       240       250
```

SEQ ID 8902 (GBS431) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 7; MW 54kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 6; MW 29kDa).

GBS431-GST was purified as shown in Figure 223, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1787

A DNA sequence (GBSx1894) was identified in *S.agalactiae* <SEQ ID 5559> which encodes the amino acid sequence <SEQ ID 5560>. This protein is predicted to be isochorismatase. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -2.81   Transmembrane  126 - 142 ( 125 - 142)

----- Final Results -----
                bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15164 GB:Z99120 similar to pyrazinamidase/nicotinamidase
          [Bacillus subtilis]
 Identities = 95/181 (54%), Positives = 132/181 (72%)

Query: 1   MTKALISIDYTYDFVADDGKLTAGKPAQSIASAIADVTEKAYRSGDYIFFAIDNHDIGDV 60
           M KALI IDYT DFVA DGKLT G+P + I  AI ++T++   +GDY+  A+D+HD GD
Sbjct: 1   MKKALICIDYTNDFVASDGKLTCGEPGRMIEEAIVNLTKEFITNGDYVVLAVDSHDEGDQ 60

Query: 61  PHPESNLPPESNIKGTSGRNLVGPLGTLYETIKEDSRVFWIDKRHYSAPSGTDLDIRLRE 120
           +HPE+ LPP HNIKGT G++LYG L  IY+  + +  V++++K  YSAF+GTDL+++LRE
Sbjct: 61  YHPETRLPPPHNIKGTBGKDLYGKLLPLYQKHEHEPNVYYMEKTRYSAPAGTDLELKLRE 120

Query: 121 RRVDTPLILTGVLTDICVLHTAIDAYNLGYKIEVPAAAVASLNDSNHQWALNHFKTVLGAT 181
           R++  L L GV TDICVLHTA+DAYN G++I V   AVAS N   H WAL+HF   +GA +
Sbjct: 121 RQIGELHLAGVCTDICVLHTAVDAYNKGFRIVVHEQAVASFNQEGHAWALSHFANSIGAQV 181
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5561> which encodes the amino acid sequence <SEQ ID 5562>. Analysis of this protein sequence reveals the following:

```
      Possible site: 31
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60   Transmembrane  126 - 142 ( 126 - 142)

----- Final Results -----
                bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15164 GB:Z99120 similar to pyrazinamidase/nicotinamidase
          [Bacillus subtilis]
 Identities = 90/179 (50%), Positives = 127/179 (70%)

Query: 3   RALISIDYTNDFVADDGKLSAGKSAQAIATKIADVTKTAFDQGDYIFFAIDCHDQNDSWH 62
           +ALI IDYTNDFVA DGKL+ G+  + I   I  +TK     GDY+  A+D HD+ D +H
Sbjct: 3   KALICIDYTNDFVASDGKLTCGEPGRMIEEAIVNLTKEFITNGDYVVLAVDSHDEGDQYH 62

Query: 63  PESKLPAAHNIKGTTGRHLVGPLAEVYSYMKQHPRVFWIDKRYYSAPSGTDLDIRLRERG 122
           PE++LP  HNIKGT G+ LYG L  +Y    +  P V++++K  YSAF+GTDL+++LRER
Sbjct: 63  PETRLPPPHNIKGTEGKDLYGKLLPLYQKHEHEPNVYYMEKTRYSAPAGTDLELKLRERQ 122

Query: 123 ITQLVLTGVLSDICVLHTAIDAYHLGYQLEIVKSAVASLTKESYEWSLAHFEQVLGAKLI 181
           I +L L GV +DICVLHTA+DAY+ G+++ + K AVAS  +E + W+L+HF   +GA++
Sbjct: 123 IGELHLAGVCTDICVLHTAVDAYNKGFRIVVHKQAVASFNQEGHAWALSHFANSIGAQV 181
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 121/180 (67%), Positives = 150/180 (83%)

Query: 3   KALISIDYTYDFVADDGKLTAGKPAQSIASAIADVTEKAYRSGDYIFFAIDNHDIGDVFH 62
           +ALISIDYT DFVADDGKL+AGK AQ+IA+ IA+VT+ A+  GDYIFFAID HD  D +H
Sbjct: 3   RALISIDYTNDFVADDGKLSAGKSAQAIATKIASVTKTAFDQGDYIFFAIDCHDQNDSWH 62

Query: 63  PESNLPPSHNIKGTSGRNLVGPLGTLYETIKEDSRVFWIDKRHYSAPSGTDLDIRLRERR 122
           PES LP  HNIKGT+GR+LVGPL  +Y   +K+  RVFWIDKR+YSAPSGTDLDIRLRER
Sbjct: 63  PESKLPAAHNIKGTTGRHLVGPLAEVYSYMKQHPRVFWIDKRYYSAPSGTDLDIRLRERG 122

Query: 123 VDTLILTGVLTDICVLHTAIDAYNLGYKIEVPAAAVASLNDSNHQWALNHFKTVLGATIL 182
           +  L+LTGVL+DICVLHTAIDAY+LGY++E+  +AVASL    +++W+L HF+ VLGA ++
Sbjct: 123 ITQLVLTGVLSDICVLHTAIDAYHLGYQLEIVKSAVASLTKESYEWSLAHFEQVLGAKLI 182
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1788

A DNA sequence (GBSx1895) was identified in *S.agalactiae* <SEQ ID 5563> which encodes the amino acid sequence <SEQ ID 5564>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1539(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1789

A DNA sequence (GBSx1896) was identified in *S.agalactiae* <SEQ ID 5565> which encodes the amino acid sequence <SEQ ID 5566>. This protein is predicted to be 3-hydroxyacyl-CoA dehydrogenase (hbd-10). Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood = -0.27   Transmembrane     3 -  19 (   1 -  19)
   INTEGRAL     Likelihood = -0.11   Transmembrane   277 - 293 ( 277 - 294)

----- Final Results -----
             bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF12219 GB:AE001862 3-hydroxyacyl-CoA dehydrogenase, putative
            [Deinococcus radiodurans]
 Identities = 151/321 (47%), Positives = 196/321 (61%), Gaps = 36/321 (11%)

Query: 56  NMTIKNLTVAGSGVLGSQIAFQAAYKGMSVTIYDINDRALNKGKERIKKLAKVVQSKIET 115
           +M+IK +TV GSGVLGSQIAPQ A+ G  V +YDIND A+ K +E + KL   YQ +++
Sbjct: 51  SMSIKTVTVCGSGVLGSQIAFQTAPHGFDVHLYDINDAAIAKARETLGKLQARYQQDLKV 110

Query: 116 AKEAYSDKAKSIKYNKNLLPSLDHIFLSKVADSLDLIADLPNQITPSKNLDQAVSDADLV 175
               +   D                                  +I+F  ++ +AV   DLV
Sbjct: 111 DAQQTGDAFA----------------------------------RISFFTDIAEAVKGVDLV 138

Query: 176 IEAVPEIVSIKEDFYKQLAKVAPSKTIFATNSSTLVPSQFADITGRPDKFLAMHFANNIW 235
            IEA+PE + IK  FY QL +VA    TIFATNSSTL+PSQF + TGRP+KFLA+HFAN IW
Sbjct: 139 IEAIPENMDIKRKFYNQLGEVADPNTIFATNSSTLLPSQFMKSTGRPSKFLALHFANEIW 198

Query: 236 QNNIVEIMGHKGTDDSVIKEALAFSKDIGMVPLHIKEQPGYILNSILVPFLESALALYY 295
           + N  EIM    TDD V    + F+KDIGMV L ++KEQ GYILN++LVP L +AL L
Sbjct: 199 KFNTAEIMRTPRTDDAVFDTVVQFAKDIGMVALPMYKEQAGYILNTLLVPLLGAALELVV 258

Query: 296 DKVSDSETIDKTWKLGTGAPMGPLELIDIGIDTAYNIMKNYSDTNSDPNSLHAHLAKML 355
             ++D +T+DKTW + TGAP GP   LD+IG+ T YNI  N +   ++P S  A  AK +
Sbjct: 259 KGIADPQTVDKTWMIATGAPRGPFAFLDVIGLTTPYNI--NMASAETNPGS--AAAAKYI 314

Query: 356 KEEPIDKGRTGKAAGHGFYDY 376
            KE +IDKG+ G A G GFY Y
```

```
Sbjct: 315 KENYIDKGKLGTATGEGFYKY 335
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8903> and protein <SEQ ID 8904> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length  of  UR:   20
     Peak Value of UR:   1.55
     Net Charge of CR: 1
McG: Discrim Score:     -0.60
GvH: Signal Score (-7.5): -3.93
     Possible site: 21
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program   count: 1 value:  -0.11 threshold:  0.0
   INTEGRAL    Likelihood = -0.11   Transmembrane  221 - 237 ( 221 - 238)
   PERIPHERAL  Likelihood =  4.61       6
 modified ALOM score:   0.52
icml HYPID: 7  CFP: 0.104

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
37.5/60.5% over 278aa
                                                              Archaeoglobus
fulgidus
  EGAD|103851| 3-hydroxyacyl-CoA dehydrogenase Insert characterized OMNI|AF2273 3-
hydroxyacyl-CoA dehydrogenase (hbd-10) Insert
 characterized
  GP|2648250|gb|AAB88983.1||AE000948 3-hydroxyacyl-CoA dehydrogenase (hbd-10) Insert
characterized
  PIR|A69534|A69534 3-hydroxyacyl-CoA dehydrogenase (hbd-10) homolog - Insert characterized

ORF01176(475 - 1431 of 1731)
EGAD|103851|AF2273(17 - 295 of 668) 3-hydroxyacyl-CoA dehydrogenase {Archaeoglobus
fulgidus}OMNI|AF2273 3-hydroxyacyl-CoA dehydrogenase (hbd-
10)GP|2648250|gb|AAB88983.1||AE000948 3-hydroxyacyl-CoA dehydrogenase (hbd-10)
{Archaeoglobus fulgidus}PIR|A69534|A69534 3-hydroxyacyl-CoA dehydrogenase (hbd-10) homolog
- Archaeoglobus fulgidus
%Match = 14.8
%Identity = 37.5  %Similarity = 60.4
Matches = 106  Mismatches = 106  Conservative Sub.s = 65

387       417       447       477       507       537       567       597
KKRYYFKNNHTIYLLLDISFVKLSSKTFSNISIGGCNMTIKNLTVAGSGVLGSQIAFQAAYKGMSVTIYDINDEALNKGK
                         :    :        || : | |:|::|  ||   |  | :||: ||  | :::|
                       MPRRVKQVINMDVRERIKTVAVLGAGLMGHGIAEVCAMAGYNVTMRDIKQEFVDRGM
                               10        20        30        40        50

624       651       681       711       741       771       801       831
ERIKK-LAKVYQS-EIETAKEAYSDKAKSIKYNKNLLPSLDHIFLSKVADSLDLIADLPNQITFSKNLDQAVSDADLVIE
  ||: ||| :|  :|::|:|  |                                    :|  : :|::|| |||||||
NMIKESLAKLEQKGKIKSAEEVLS------------------------------------RIKPTVDLEEAVKDADLVIE
           70        80                                            90        100
```

```
861         891         921         951         981        1011        1041        1071
AVPETVSIKEDFYKQLAKVAPSKTIFATNSSTLVPSQFADITGRPDKFLAMHFANNIWQNNIVEIMGHKGTDDEVIKEAL
|||| | ||:  :::: |:|    || :|:||:  : :|| | ||:||  :|| |      :||::  : | |||:   :
AVPEVVEIKKQVWEEVDKLAKPDCIFTSNTSTMRITMLADFTSRPEKFAGLHFFNPPVLMRLVEVIRGEKTSDEVMDLLV
                120         130         140         150          160         170         180

1101        1131        1161        1191        1221        1251        1281        1311
AFSKDIGMVPLHIHKEQPGYILNSILVPFLESALALYYDKVSDSETIDKTWKLGTGAPMGPLEILDIIGIDTAYNIMKNY
 | | ||  |: : |: ||:|:| :  |     :|:    ::  | :| | :   | ||||:|::|  |:|  || :| |
EFVKSIGKTPVRVEKDVPGFIVNRVQAPASVLLMAILEKGIATPEEVDATVR-RLGLPMGPFELVDYTGVDILYNALKYY
                200         210         220         230          240         250         260

1341        1371        1401        1431        1461        1491        1521        1551
SDTNSDPNSLHAHLAKMLKEEFIDKGRTGKAAGHGFYDYD*TIKEVR*KSNLFYNSTKE*LHQEQF*NDLKPIDDYYHLS
: | | |:    :    : || :   : |:  |:||| :     ::         :  :    |
AQTIS-PD----YEPPKFLEEMVKANKLGRKTGQGFYDWSKGRPQIDSSKATDKINPMDFTFVEINEAVKLVEMGVATPQ
            270         280         290         300         310         320         330
```

SEQ ID 8904 (GBS112) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 28 (lane 5; MW 39kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 3; MW 64kDa).

GBS112-GST was purified as shown in Figure 198, lane 10.

Example 1790

A DNA sequence (GBSx1897) was identified in *S.agalactiae* <SEQ ID 5567> which encodes the amino acid sequence <SEQ ID 5568>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3332(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10097> which encodes amino acid sequence <SEQ ID 10098> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14467 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 62/169 (36%), Positives = 109/169 (63%), Gaps = 3/169 (1%)

Query: 1    MAVLSMLGIIDAKPKVGYFYLGQYHASIGTSHFEKMTVSEIMGIPLTVHQKDSVYDVIVH 60
             +A+L+M G ++A+P+VGYFY G+     +      +K+ V +   IP+ +H+  SVYD I
Sbjct: 43   LAILTMSGFLEARPRVGYFYTGKTGTQLLADKLKKLQVKDFQSIPVVIHENVSVYDAICT 102

Query: 61   IFMEDAGCAFILDDDDFLCGVVSRKDLLKISIGGGDLSKMPIGMVMTRMPHVTTVLENES 120
             +F+ED G  F++D D  L GV+SRKDLL+ SIG  +L+ +P+ ++MTRMP++T      +
Sbjct: 103  MFLEDVGTLFVVDRDAVLVGVLSRKDLLRASIGQQELTSVPVHIIMTRMPNITVCRREDY 162

Query: 121  LFAAADKLVSRKVDSLPVVRHDKQYPEKFKVIGKLSKTILASLFLEIRD 169
             +   A  L+ +++D+LPV+   K   + F+VIG+++KT +  + + + +
Sbjct: 163  VMDIAKHLIEKQIDALPVI---KDTDKGFEVIGRVTKTNMTKILVSLSE 208
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1791

A DNA sequence (GBSx1898) was identified in *S.agalactiae* <SEQ ID 5569> which encodes the amino acid sequence <SEQ ID 5570>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.53   Transmembrane   60 -  76 (  60 -  76)

----- Final Results -----
                bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05092 GB:AP001511 unknown conserved protein [Bacillus halodurans]
 Identities = 126/256 (49%), Positives = 183/256 (71%), Gaps = 1/256 (0%)

Query: 7   IFIISDSIGETAKAIAKACLSQFPGHDDWHPQRPSYINSQERLRQVFEERASQKTVFMMFS 66
            ++++SDS+GETA+ + KA  SQF G       +R  Y+  +E +++V + A Q    + F+
Sbjct: 10  VTVVSDSVGETAELVVKAAASQFSGAGI-EVRRIPYVEDKETVDEVIQLAKQADAIIAFT 68

Query: 67  LVDVALASYAQKRCESEHYAYVDLLTNVIQGISRISGIDPLGEPGILRRLDNDYFKRVES 126
            LV   + +Y  ++          VD++  +++ IS ++  +P  EPGI+ RLD DYP++VE+
Sbjct: 69  LVVPGIRTYLLEKATEAKVETVDIIGPMLEKISSLTKEEPRYEPGIVYRLDEDYPRKVEA 128

Query: 127 IEFAVKYDDGRDPRGILQADLVIIGISRTSKTPLSMFLADKNIKVINIPLVPEVDVPKEL 186
            IEFAVKYDDGRDPRGI++ADLV+IG+SRTSKTPLS +LA K +KV N+PLVPEV   P+EL
Sbjct: 129 IEFAVKYDDGRDPRGIVRADLVLIGVSRTSKTPLSQYLAHKRLKVANVPLVPEVEPPEEL 188

Query: 187 RMIDSRRIIGLTNSVDHLNQVRKVRLKSLGLSSTANYASLERILEETRYAEEVMKNLGCP 246
              +  +++IGL  S + LN +R  RLK+LGL S ANYA+++RI EE  YAE +MK +GCP
Sbjct: 189 FKLSPKKVIGLKISPEQLNGIRARRLKTLGLKSQANYANIDRIKEELAYAEGIMKRIGCP 248

Query: 247 IINVSDKAIEETATII 262
            +I+VS+KA+EETA +I
Sbjct: 249 VIDVSNKAVEETANLI 264
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5570 (GBS378) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 4; MW 34kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 2; MW 59kDa).

GBS378-GST was purified as shown in Figure 212, lane 6.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1792

A DNA sequence (GBSx1899) was identified in *S.agalactiae* <SEQ ID 5571> which encodes the amino acid sequence <SEQ ID 5572>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3703(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD35361 GB:AE001709 pyruvate,orthophosphate dikinase
          [Thermotoga maritima]
 Identities = 494/882 (56%), Positives = 639/882 (72%), Gaps = 9/882 (1%)

Query: 1   METKPVYHFD----EGCKEMKELLGGKGANLAEMTSIGLFVPQGFTITTQACNDYYDNAC 56
           M  K+VY F      EG  +MK++LGGKGANLAEMT+++G+PVP GFTI+ + C  YYD+
Sbjct: 1   MAKKVVYFFANGKAEGRADMKDILGGKGANLAEMTNLGIPVPPGFTISAEVCKYYYDHGR 60

Query: 57  HIRESILSQIDQAMAQLEVEQNKQLGSVDDPLLVSVRSGSVFSMPGMMDTVLNLGLNDRS 116
            E +  Q+++AM +LE     K+ G  ++PLLVSVRSG+   SMPGMMDTVLNLGLND +
Sbjct: 61  TYPEELKEQVERAMRRLEEVTGKKPGDPNNPLLVSVRSGAAISMPGMMDTVLNLGLNDET 120

Query: 117 VQGLVKKTRDERFAYDSYRRFIQMFADVVTGIPKYKFDTILDRLKTDKCYQDDTELTGSD 176
            V+G+ K T +ERFAYD+YRRF+QMF DVV  IP  KF+  L+ LK +K  + DTEL   D
Sbjct: 121 VKGLAKLTNNERFAYDAYRRFLQMFGDVVLKIPHEKPEKALELLKEKGVKLDTELDAED 180

Query: 177 LKRLVKFYKELVQKEAGEKFPQDPKRQLLLAIEAVFKSWNNPRAKIYRKLNDIPE--TLG 234
            LK+LVE YK++Y KE G++FPQDP +QL LAI+AVF SW N RA  YR+++ I E   LG
Sbjct: 181 LKKLWERYKQIY-KEEGKEFPQDPWKQLWLAIDAVFGSWMNSRAIKYRQIHGIKEGDLLG 239

Query: 235 TAVNIQAMVFGNMGNNSGTGVAFTRNPSTGAANLPGEYLINAQGEDVVAGIRTPQSISKL 294
            TAVNI AMVFGNMG +SGTGVAFTR+P+TG     +GE+L NAQGEDVVAGIRTP  + +L
Sbjct: 240 TAVNIVAMVFGNMGEDSGTGVAFTRDPNTGEKKPYGEFLPNAQGEDVVAGIRTPLKLEEL 299

Query: 295 AEQMPIIYQEFVSVTQKLEAHYRDMQDMEPTIENGNLYMLQTRSGKRTAKAAIKIAVDQV 354
              +MP +Y + + +   KLE HYRDMQD+EPT+E G LY+LQTR+GKRT++AAI+IAVD V
Sbjct: 300 KNRMPEVYNQLLEIMDKLEKHYRDMQDIEPTVERGKLYILQTRNGKRTSQAAIRIAVDMV 359

Query: 355 NEGLISKEEAILRIEPKQLDQLLHPSFDLKSLKKAIILTTGLPASPGAAYGKVYFHAEDV 414
           +EGLI+KEEAILR+ P+ ++Q+LHP FD K   +A ++  GLPASPGAA GKV F+A+
Sbjct: 360 HEGLITKEEAILRVRPEDVEQVLHPVFDPKEKAQAKVIAKGLPASPGAATGKVVFNAKKA 419

Query: 415 VKEMKKGNPVLLVRQETSPEDIEGMVSANGIITARGGMTSHAAVVARGMGKPCVAGCSQL 474
            +  K G  V+LVR ETSPED+ GM +A GI+T+RGGMTSHAAVVARGMGKP V G   +
Sbjct: 420 EELGKAGEQVILVRPETSPEDVGGMAAAQGILTSRGGMTSHAAVVARGMGKPAVVGAESI 479

Query: 475 LVDEVRREISIGHQTIKEGEMLSIDGATGNVYIGQV-PMAETSVDRDPFEIFMKWVDENRD 533
             V         +G   +KEGE +SIDG TG V +G+V  +     ++W DE R
Sbjct: 480 EVHPEEGYFKVGDVVVKEGEWISIDGTTGEVLLGKVTTIKPQSLEGPVAELLQWADEIRR 539

Query: 534 MKVCSNADNPRDAQKALDFGAEGIGLCRTEHMFFDDERIPVVREMILADEILSRRKALER 593
            + V +NAD PRDA+ A  FGAEGIGLCRTEHMFF+ +RIP VR MILA     R KAL+
Sbjct: 540 LGVRTNADIPRDAEVARKFGAEGIGLCRTEHMFFEKDRIPVVREMILAKTKEERESKALDE 599

Query: 594 LLSFQRDDFVQIFKVLKGKACTIRLLDPPLHEFLPHDKESIRSMARQMGISTLLAIEKRIQ 653
            LL  Q++DF  +F+V+++KG   TIRLLDPPLHEFLP + E I+ +A QMG+S  ++  ++
Sbjct: 600 LLPLQKEDFKGLFRVMKGLEVTIRLLDPPLHEFLPQEDEQIKEVABQMGVSFEELKNVVE 659

Query: 654 TLEEFNPMLGHRGCRLAITYPEIYQMQVRALVQGAI-LAMKEGYEAKPEIMIPLVTAHEE 712
             L+E NPMLGHRGCRL ITYPEI  MQ +A++  AI L  +EG +   PEIMIPLV   E
Sbjct: 660 NLKEELNPMLGHRGCRLTITYPEIAVMQTKAIIGAAIELKKEGIDVIPEIMIPLVGHVNE 719

Query: 713 ISIIRDLIEETIVEESKSKKINLSFPIGTMIETPRACMIADDIAKFADFFSFGTNDLTQM 772
           +  ++ +I+ET     K   + L++ IGTMIE PRA + A   IA+ A+FFSFGTNDLTQM
Sbjct: 720 LRYLKKIIKETADALIKEAGVELTYKIGTMIEVPRAALTADEAEFFSFGTNDLTQM 779

Query: 773 SFGFSRDDAGKFLGEYVDKGILKKDPFQVLDQKGIGRFIGQAVRLGKEVKPNLKIGICGE 832
           +FGFSRDD GKFL EY++KG+L+ DPF+ LD  G+G  +     G+   +P+LK+G+CGE
Sbjct: 780 TFGFSRDDVGKFLPEYLEKGILEHDPFKTLDVDGVGELVRMGKEKGRSTRPDLKVGVCGE 839

Query: 833 HGGEPSSIEFCYQLGLHYVSCSPFRIPIAKLAAAQAKIKQSR 874
           HGG+P SI F  ++GL YVSCSP+R+P+A+LAAAQA +K  +
Sbjct: 840 HGGDPRSILFFDKIGLDYVSCSPYRVPVARLAAAQAALKNKK 881
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1793

A DNA sequence (GBSx1900) was identified in *S.agalactiae* <SEQ ID 5573> which encodes the amino acid sequence <SEQ ID 5574>. This protein is predicted to be glutamyl-tRNA (Gln) amidotransferase subunit C (gatC). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3229(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04384 GB:AP001509 glutamyl-tRNA (Gln) amidotransferase
          subunit C [Bacillus halodurans]
 Identities = 42/94 (44%), Positives = 63/94 (66%)

Query: 2  KISEEEVRHVANLSKLRFSDQETKEFASSLSKIVDMIELLNEVDTEGVEPVTTTMADRKTV 61
          +IS E+V+HVA+L++L   +++E K F    L  I+    E LNE+DTEGV  T+ + D K V
Sbjct: 3  RISMEQVKHVAHLARLAITEEEAKLFTEQLGDIIQFAEQLNELDTEGVEPTSHVLDMKNV 62

Query: 62 MREDLAQPGHNRDDLFKNVPQHQDYYIKVPAILE 95
          +RED  + G   +D+ KN P H+D  I+VP++LE
Sbjct: 63 LREDKPEKGLPVEDVLKNAPDHEDGQIRVPSVLE 96
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5575> which encodes the amino acid sequence <SEQ ID 5576>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3247(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 72/100 (72%), Positives = 88/100 (88%)

Query: 1   MKISEEEVRHVANLSKLRFSDQETKEFASSLSKIVDMIELLNEVDTEGVPVTTTMADRKT 60
           MKISEEEVRHVA LSKL FS+ ET  FA++LSKIVDM+ELLNEVDTEGV +TTTMAD+K
Sbjct: 5   MKISEEEVRHVAKLSKLSFSESETTTFATTLSKIVDMVELLNEVDTEGVAITTTMADKN 64

Query: 61  VMREDLAQPGHNRDDLFKNVPQHQDYYIKVPAILEKGGDA 100
           VMR+D+A+ G +R  LFKNVP+ ++++IKVPAIL+DGGDA
Sbjct: 65  VMRQDVAEEGTDRALLFKNVPEKENHFIKVPAILDDGGDA 104
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1794

A DNA sequence (GBSx1901) was identified in *S.agalactiae* <SEQ ID 5577> which encodes the amino acid sequence <SEQ ID 5578>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood = -7.64   Transmembrane    7 -  23 (   6 -  24)

----- Final Results -----
               bacterial membrane --- Certainty=0.4057(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1795

A DNA sequence (GBSx1902) was identified in *S.agalactiae* <SEQ ID 5579> which encodes the amino acid sequence <SEQ ID 5580>. This protein is predicted to be glutamyl-tRNA amidotransferase, subunit A (gatA). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2855(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04385 GB:AP001509 glutamyl-tRNA (Gln) amidotransferase
          subunit A [Bacillus halodurans]
 Identities = 285/486 (58%), Positives = 367/486 (74%), Gaps = 4/486 (0%)

Query: 1   MSFNNQSIDQLHDFLVKKEISATELTKATLEDIHARSQAVGSFITTISDEMAIAQAKEID- 59
            MS  +  +  +H   L +KEIS ++L      + I    +  V +F+ +++E A A AKE+D
Sbjct: 1   MSLFDLKLKDVHTKLHHKEISVSDLVDEAYKRIEQVDQQVEAFLALNEKARAYAKELDA 60

Query: 60  --DKGIDADNVMSGIPLAVKDNISTKGILTTAASKMLYNYEPIFDATAVEKLYAKDMIVI 117
              D+  +A   ++ GIP+ VKDNI TK + TT +S++L N++PI+DAT V KL      + I
Sbjct: 61  ALDRS-EARGLLFGIPIGVKDNIVTKNLRTTCSSRILGNFDPIYDATVVHKLREAQAVTI 119

Query: 118 GKANMDEFAMGGSTETSYFKKTKNPWNLEYVPGGSGGSGAAAVASGQVRLSLGSDTGGSI 177
            GK NMDEFAMG STE S F+KT N W+   VPGGSGGSGAAAVA+G+V  +LGSDTGGSI
Sbjct: 120 GKLNMDEFAMGSSTENSAFQKTKNPWNLEYVPGGSGGSGAAAVAAGEVPFTLGSDTGGSI 179

Query: 178 RQPASFNGIVGMKPTYGRVSRFGLFAFGSSLDQIGPMSQTVKENAQLLTVISGHDVPDST 237
            RQPA++ G+VG+KPTYGRVSR+GL AF SSLDQIGP+++ V++NA LL  ISGHD  DST
Sbjct: 180 RQPAAYCGVVGLKPTYGRVSRYGLVAFASSLDQIGPITNNVEDNAYLLQAISGHDPMDST 239

Query: 238 SSERTVGDFTAKIQQDIQGMKIALPKEYLGEGIAQGVKSTIIKAAKHLEKLGAVIEEVSL 297
            S+   V D+ + +  DI+G+KIA+PKEYLGEG+ + VK+++ A K LE LGA  EEVSL
Sbjct: 240 SANLDVPDYLGALTGDIKGLKIAVPKEYLGEGVKEEVKQSVLDALKVLEGLGATWEEVSL 299

Query: 298 PHSKYGVAVYYIVASSEASSNLQRFDGIRYGYRTENYKNLDDIYVMTRSEGFGDEVKRRI 357
            PHSKY +A YY++ASSEAS+NL RFDG+RYG+R++N  NL D+Y  TR+EGFGDEVKRRI
Sbjct: 300 PHSKYALATYYLLASSEASANLARFDGVRYGFRSDNADNLLDMYKQTRAEGFGDEVKRRI 359

Query: 358 MLGTFSLSSGYYDAYYKKAQQVRSLIIQDFEKVFADYDLILGPTAPTTAFDLDSLNEDPV 417
            MLGTF+LSSGYYDAYYKKA QVR+LI QDFEKVF  YD+I+GPT PT AF +     DP+
Sbjct: 360 MLGTFALSSGYYDAYYKKAQQVRTLIKQDFEKVFEQYDVIIGPTTPTPAFKIGEKTDDPL 419

Query: 418 AMYLADILTIPVNLAGLPGISIPAGFDQGLPVGMQLIGPKFSEETIYQVAAAFEATTDYH 477
```

```
              MY  DILTIPVNLAG+P IS+P GFD GLP+G+Q+IG  F E ++Y+VA AFE  TDYH
Sbjct: 420 TMYANDILTIPVNLAGVPAISVPCGFDNGLPLGLQIIGKHFDEGSVYRVAHAFEQATDYH 479

Query: 478 KQQPKI 483
           ++P +
Sbjct: 480 TKRPTL 485
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5581> which encodes the amino acid sequence <SEQ ID 5582>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2364(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 392/487 (80%), Positives = 442/487 (90%)

Query: 1   MSFNNQSIDQLHDFLVKKEISATELTKATLEDIHARERQAVGSFITISDEMAIAQAKEIDD 60
            MSFN++++I++LHD LV KEISATELT+ATLEDI +RE+AVGSFIT+S+E+A+ QA  ID
Sbjct: 1   MSFNHKTIEELHDLLVAKEISATELTQATLEDIKSRREAVGSFITVSEEVALKQAAAIDA 60

Query: 61  KGIDADNVMSGIPLAVKDNISTKGILTTAASKMLYNYEPIFDATAVEKLYAKDMIVIGKA 120
            KGIDADN+MSGIPLAVKDNISTK ILTTAASKMLYNYEPIF+AT+V   YAKDMIVIGK
Sbjct: 61  KGIDADNLMSGIPLAVKDNISTKEILTTAASKMLYNYEPIFNATSVANAYAKDMIVIGKT 120

Query: 121 NMDEFAMGGSTETSYFKKTNNAWDHSKVPGGSSGGSAAAVASGQVRLSLGSDTGGSIRQP 180
            NMDEFAMGGSTETSYFKKT NAWDH+KVPGGSSGGSA AVASGQVRLSLGSDTGGSIRQP
Sbjct: 121 NMDEFAMGGSTETSYFKKTKNAWDHTKVPGGSSGGSATAVASGQVRLSLGSDTGGSIRQP 180

Query: 181 ASFNGIVGMKPTYGRVSRFGLFAPGSSLDQIGPMSQTVCENAQLLTVISGHDVRDSTSSE 240
            A+FN +VG+KPTYG VSR+GL APGSSLDQIGP + TVKENAQLL VI+  DV+D+TS+
Sbjct: 181 AAFNSVVGLKPTYGTVSRYGLIAPGSSLDQIGPFAPTVKENAQLLNVIASSDVKDATSAP 240

Query: 241 RTVGDPTAKIGQDIQGMKIALPKEYLGEGIAQGVKETIIKAAKHLEKLGAVIEEVSLPHS 300
              + D+T+KIG+DI+GMKIALPKEYLGEGI    +KET++ + K  E LGA +EEVSLPHS
Sbjct: 241 VRIADYTSKIGRDIKGMKIALPKEYLGEGIDPEIKETVLASVKQFEALGATVEEVSLPHS 300

Query: 301 KYGVAVYYIVASSEASSNLQRFDGIRYGYRTENYKNLDDIYVNTRSRGFGDEVKRRIMLG 360
            KYGVAVYYI+ASSEASSNLQRFDGIRYG+R ++ KNLD+IYVNTRS+GFGDEVKRRIMLG
Sbjct: 301 KYGVAVYYIIASSEASSNLQRFDGIRYGFRADDAKNLDEIYVNTRSQGFGDEVKRRIMLG 360

Query: 361 TFSLSSGYYDAYYKKAQQVRSLIIQDFEKVFADYDLILGPTAPTTAFDLDSLNHDPVANY 420
            TFSLSSGYYDAY+KKAQQVR+LIIQDF+KVFADYDLILGPT PT AF LD+LNHDPVANY
Sbjct: 361 TFSLSSGYYDAYFKKAQQVRTLIIQDFDKVFADYDLILGPTTPTVAPGLDTLNHDPVANY 420

Query: 421 LADILTIPVNLAGLPGISIPAGFDQGLPVGMQLIGPKFSEETIYQVAAAFEATTDYHKQQ 480
            LAD+LTIPVNLAGLPGISIPAGF  GLPVG+QLIGPK++EETIYQ AAAFEA TDYHKQQ
Sbjct: 421 LADLLTIPVNLAGLPGISIPAGFVDGLPVGLQLIGPKYAEETIYQAAAAFEAVITDYHKQQ 480

Query: 481 PKIFGGE 487
            P IFGG+
Sbjct: 481 PIIFGGD 487
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1796

A DNA sequence (GBSx1903) was identified in *S.agalactiae* <SEQ ID 5583> which encodes the amino acid sequence <SEQ ID 5584>. This protein is predicted to be glutamyl-tRNAGln amidotransferase subunit B (gatB). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3935(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10095> which encodes amino acid sequence <SEQ ID 10096> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04386 GB:AP001509 glutamyl-tRNA (Gln) amidotransferase
        subunit B [Bacillus halodurans]
 Identities = 308/476 (64%), Positives = 361/476 (75%), Gaps = 1/476 (0%)

Query: 1   MNFETVIGLEVHVELNTNSKIFSPSSAHFGQEQNANTNVIDWSFPGVLPVMNKGVIDAGI 60
            MNFETVIGLEVHVEL T SKIFS S  HFG E NANT+VID  +PGVLPV+NK  ++  +
Sbjct: 1   MNFETVIGLEVHVELKTESKIFSASPNHFGAEPNANTSVIDLGYPGVLPVLNKAAVEFAM 60

Query: 61  KAALALNMDIHQNMMPDRKNYFYPDNPKAYQISQFDEPIGYNGWIEIELEDGTRKKIRIE 120
            KAA+ALN ++  + PDRKNYFYPDNPKAYQISQFD+PIG NGWIEIE+ DGT+KKI I
Sbjct: 61  KAAMALNCEVATDTKPDRKNYFYPDNPKAYQISQFDKPIGENGWIEIEV-DGTKKKIGIT 119

Query: 121 RAHLEEDAGKNTHGTDGYSYVDLNRQGVPLIEIVSEADMRSPEEAYAYLTALKEIIQYTG 180
            R HLEEDAGK TH  +GYS VD NRQG PLIEIVSE D+R+P+EAYAYL  LK IIQYTG
Sbjct: 120 RLHLEEDAGKLTHSGNGYSLVDFNRQGTPLIEIVSEPDIRTPQEAYAYLEKLKSIIQYTG 179

Query: 181 ISDVKMEEGSMRVDANISLRPYGQEEFGTKAELKNLNSFNMVRKGLIHEEKKQAQVLRSG 240
            +SD KMEEGS+R DANISLRP GQEEFGTK ELKNLNSFN VRKGL +EEKKQAQVL SG
Sbjct: 180 VSDCKMEEGSLRCDANISLRPVGQEEFGTKTELKNLNSFNFVRKGLEYEEKRQAQVLLSG 239

Query: 241 GQIQQTRTRFDETTGETILMRVKEGSSDYRYFPEPDLLPLDISDEWIDQVRLELPEFPQE 300
            G+I QETRR+DE   +T+LMRVKEGS DYRYFPEPDL     I DEW  ++R E+PE P
Sbjct: 240 GEILQETRRYDEAANKTVLMRVKEGSDDYRYFPEPDLVALHIDDEWKARIRSEIPELPDA 299

Query: 301 RRAKYVSSPGLSSYDASQLTATKATSDFFEKAVAIGGDAKQVSNWLQGEVAQFLNSESKS 360
            R+ +VV   GL +YDA  LT TK  SDFFE+ +A G D K  SNWL GEV+ +LN+E K
Sbjct: 300 RKKRYVELGLLPAYDAMVLTLTKEMSDFFESTIAKGADPKLASNWLMGEVSGYLNAEQKE 359

Query: 361 IEEIGLTPENLVKMIGLIADGTISSKIAKKVFVHLAKNGGSAEEFVKKAGLWQISDPEVL 420
            ++E+ LTP+ L +MI LI  GTISSKIAKKVF  L + GG  EE VK  GLWQISD   L
Sbjct: 360 LDEVALTPDGLAKMIQLIEKGTISSKIAKKVFKDLIEKGGDPEEIVKAKGLWQISDDGEL 419

Query: 421 IPIIHQVFADNEANVIDFKSGKRNADKAPTGYLMKATKGQANDQVALKLLAQELAK 476
                + +V  +N+ ++ DFK+GK  A     G +MKATKG+ANP +  KLL +E+ K
Sbjct: 420 RKYVVEVLDNNQQSIDDFKNGKDRAIGFLVGQIMKATKGKANPPMVNKLLLEEINK 475
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5585> which encodes the amino acid sequence <SEQ ID 5586>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3935(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 410/479 (85%), Positives = 447/479 (92%)

Query: 1   MNFETVIGLEVHVELNTNSKIFSPSSAHFGQPGTNANTNVIDWSFPGVLPVMNKGVIDAGI 60
           MNFET+IGLEVHVELNTNSKIFSPSSAHFG++ NANTNVIDWSFPGVLPVMNKGVIDAGI
Sbjct: 1   MNFETIIGLEVHVELNTNSKIFSPSSAHFGEDPNANTNVIDWSFPGVLPVMNKGVIDAGI 60

Query: 61  KAALALNMDIHQNMHFDRKNYFYPDNPKAYQISQFDEPIGYNGWIEIELEDGTRKKIRIE 120
           KAALALNMDIH+ MHFDRKNYFYPDNPKAYQISQFDEPIGYNGWI+I+LEDG+ KKIRIE
Sbjct: 61  KAALALNMDIHKSMHFDRKNYFYPDNPKAYQISQFDEPIGYNGWIDIKLEDGSTKKIRIE 120

Query: 121 RAHLEEDAGKNTHGTDGVSYVDLNRQGVPLIEIVSEADMRSPEEAYAYLTALKEIIQYTG 180
           RAHLEEDAGKNTHGTDGYSYVDLNRQGVPLIEIVSEADMRSPEEAYAYLTALKEIIQYTG
Sbjct: 121 RAHLEEDAGKNTHGTDGYSYVDLNRQGVPLIEIVSEADMRSPEEAYAYLTALKEIIQYTG 180

Query: 181 ISDVKMEEGSMRVDANISLRPYGQEEFGTKAELKNLNSFNSVRKGLIHEEKRQAQVLRSG 240
           ISDVKMEEGSMRVDANISLRPYGQE+FGTK ELKNLNSF+NVRKGL  E +RQA++LRSG
Sbjct: 181 ISDVKMEEGSMRVDANISLRPYGQEQFGTKTELKNLNSFSNVRKGLEPEVERQAKLLRSG 240

Query: 241 GQIQQETRRFDETTGETILMRVKEGSSDYRYFPEPDLPLFDISDEWIDQVRLKLPEFPQE 300
           G I+QETRR+DE    TILMRVKEG++DYRYFPEPDLPLL++I D WID++R +LP+FP +
Sbjct: 241 GVIRQETRRYDEANKGTILMRVKEGAADYRYFPEPDLPLYEIDDAWIDEMRAQLPQFPAQ 300

Query: 301 RRAKYVSSFGLSSYDASQLTATKATSDFFEKAVAIGGDAKQVSNWLQGEVAQFLNSESKS 360
           RRAKY    GLS+YDASQLTATK  SDFFE AV++GGDAKQVSNWLQGEVAQFLN+E K+
Sbjct: 301 RRAKYEELGLSAYDASQLTATKVLSDFFETAVSLGGDAKQVSNWLQGEVAQFLNAESGKT 360

Query: 361 IEEIGLTPENLVEMIGLIADGTISSKIAKKVFVHLAKNGGSAEEFVKKAGLWQISDPEVL 420
           IEEI LTPENLVEMI +IADGTISSK+AKKVFVHLAKNGGSA  +V+KAGLWQISDP VL
Sbjct: 361 IEEIALTPENLVEMIAIIADGTISSKMAKKVFVHLAKNGGSARAYVEKAGLWQISDPAVL 420

Query: 421 IPIIHQVFADNEAAVIDPKSGKRNADKAFTGYLMKATKGQANPQVALKLLAQELAKLKE 479
           +PIIHQVFADNEAAV DPKSGKRNADKAFTG+LMKATKGQANPQVA +LLAQEL KL++
Sbjct: 421 VPIIHQVFADNEAAVADPKSGKRNADKAFTGFLMKATKGQANPQVAQQLLAQELQKLRD 479
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1797

A DNA sequence (GBSx1904) was identified in *S.agalactiae* <SEQ ID 5587> which encodes the amino acid sequence <SEQ ID 5588>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood = -7.27   Transmembrane  108 - 124 ( 105 - 125)
   INTEGRAL     Likelihood = -7.27   Transmembrane  278 - 294 ( 268 - 301)
   INTEGRAL     Likelihood = -6.05   Transmembrane  191 - 207 ( 188 - 208)
   INTEGRAL     Likelihood = -5.63   Transmembrane  219 - 235 ( 215 - 242)
   INTEGRAL     Likelihood = -3.93   Transmembrane   41 -  57 (  39 -  58)
   INTEGRAL     Likelihood = -3.88   Transmembrane  132 - 148 ( 131 - 150)
   INTEGRAL     Likelihood = -3.03   Transmembrane  254 - 270 ( 253 - 272)
   INTEGRAL     Likelihood = -3.03   Transmembrane   79 -  95 (  79 -  95)

----- Final Results -----
               bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10093> which encodes amino acid sequence <SEQ ID 10094> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA04271 GB:AJ000733 hypothetical protein [Bacillus megaterium]
 Identities = 102/292 (34%), Positives = 169/292 (56%), Gaps = 3/292 (1%)

Query: 6   TKKEKGTMMTLAAGLAWGISGSQQYLMSH-GVHVNLLTSLRLLLTGIFLLSLARSKQKE 64
           +++  G ++ +     WG+SG   QYL  H   +    L  +R+L++G+ LL++A SKQ+
Sbjct: 1   SRRAWGLLLVIIGATMWGVSGTVAQYLFQHKSFNAEWLVVVRMLVSGLLLLAIA-SKQR- 58

Query: 65  HLVAAWKQPKFLKQVLLFSIPGLVLNQYAFLRAHLTNAGTATVLQTMAPILLLSIVCIL 124
           ++ A WK  +    +LLF + G++  QY +  AI    NA TATVLQY +PI I+  + +
Sbjct: 59  NIPAIWGTKSERTSLLLFGVIGMLGVQYTTFAAIEAGNAATATVLQYTSPIFIIGYLAVQ 118

Query: 125 NRQRPTSFEIIAIAMAILGTVMIATHGKLGSLAITPKGLAWWGLGSAITYSIYILLPVKLI 184
            R+ P   E+I++ + I GT+ +AT G    L+IT    L WG+G+A+T + Y L P +L+
Sbjct: 119 ARKWPVKVEMISVVLVIAGTFFLATSGNFNELSITQWALFWGIGAAVTSAFYTLQPKRLL 178

Query: 185 HEWGSTIVIGSGMPIGGILFSLVTKAWQYPLQINVMGILAYIGIIGIGTIPAYTFFLKGV 244
           +W S  V+G GM IGG  FS +   W    + +++S+ A + +I  GT+ A+  +L+ +
Sbjct: 179 AKWSSIEVVGWGMVIGGASFSFIHPPWHIAGEWSLLSLCAVLFVIIFGTLIAFYCYLESL 238

Query: 245 SIVGAVKGSLLASVEPVSSVFLTVLVLGEIFYPIDLLGMLFIFLAVTLISYK 296
            + A +   +LAS EP+S+  L+VL L   F   + LG + I   V L+S +
Sbjct: 239 KHISASAIVLASREPLSAAALSVLWLHVTFGWTEWLGTLLIIATVFLLSQR 290
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1798

A DNA sequence (GBSx1905) was identified in *S.agalactiae* <SEQ ID 5589> which encodes the amino acid sequence <SEQ ID 5590>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2103(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10091> which encodes amino acid sequence <SEQ ID 10092> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14510 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 52/153 (33%), Positives = 88/153 (56%)

Query: 17  YRPTFVVRAVYDLRAEDLLRHGIRAVLVDLDNTLIAWNNPDGTAEVRAWLDEMTTADISV 76
           + P   V+ ++ +  E L     ++ ++ DLDNTL+ W+ P+ T  +  W +EM    I V
Sbjct: 6   FLPDEFVKNIPHITPEKLKERNVKGIITDLDNTLVEWDRPNATPRLIEWFEEMKEHGIKV 65

Query: 77  VVVSNNNHARVERAVSRFGVDFVSRAMKPFTRGINMAIERYGFDRDEVIMVGDQLMTDIR 136
            +VSNNN  RV+      G+ F+ +A KP  +  N A+      +++ +++GDQL+TD+
Sbjct: 66  TIVSNNNSERVKLPSEPLGIPFIYKARKPMGRAFNRAVRNMELKKHDCVVIGDQLLTDVL 125

Query: 137 AGHRAGIKSVLVKPIVKSDAWNTKPNRLRRRRV 169
            +R G  ++LV P+  SD + T+PNR  ERR+
Sbjct: 126 GGNRNGYHTILVVPVASSDGFITRFNRQVERRI 158
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5591> which encodes the amino acid sequence <SEQ ID 5592>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4252(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 147/175 (84%), Positives = 158/175 (90%)

Query: 12  LSIDDYRPTFVVEAVYDLRAEDLLRHGIRAVLVDLDNTLIAWNNPDGTAEVRAWLDEMTT 71
           +SIDDYRPT++VEA+YDLRA DLLRHGI AVLVDLDNTLIAWNNPDGT EVRAWLDEMT
Sbjct: 20  MSIDDYRPTVMVEAIYDLRANDLLRHGITAVLVDLDNTLIAWNNPDGTPEVRAWLDEMTI 79

Query: 72  ADISVVVVSNNNHARVERAVSRPGVDFVSRAMKPFTRGINMAIERYGFDRDEVIMVGDQL 131
           ADISVVVVSNN H+RVERAVSRPGVDF+SRA+KPF  GI  AI RYGFDR+EVIMVGDQL
Sbjct: 80  ADISVVVVSNNKHSRVERAVSRPGVDFISRALKPFAYGIKKAIARYGFDRNEVIMVGDQL 139

Query: 132 MTDIRASHRAGIKSVLVKPIVKSDAWNTKFNRLRERRVWKKIEENYGKIVYQKGI 186
           MTDIRASHRAGIKSVLVKP+V SDAWNTK NR RERRV  K+EE YGK+ YQKGI
Sbjct: 140 MTDIRASHRAGIKSVLVKPLVASDAWNTKINRWRERRVMAKLEEKYGKLSYQKGI 194
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1799

A DNA sequence (GBSx1906) was identified in *S.agalactiae* <SEQ ID 5593> which encodes the amino acid sequence <SEQ ID 5594>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1091(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14509 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 220/373 (58%), Positives = 280/373 (74%), Gaps = 8/373 (2%)

Query: 1   MEKLFCIGCGARIQTENKDAAGYTPRAALEKGLETGKLYCQRCPRLHYNEITDVHTDD 60
           ME++ CIGCG  IQTE+K   GY P A+L K        + CQRCPRL++YNEI DV +TDD
Sbjct: 1   MEKVVCIGCGVTIQTEDKTGLGYAPPASLTKR----NVICQRCPRLKNYNEIQDVSLTDD 56

Query: 61  EFLKLLHEVGDSDALVVNVIDIPDFNGSIIPGLSRFVAGNDVLLVGNKKDILPKSVKDGK 120
           +FL +LH +G++D+LVV ++DIPDFNGS I GL R V GN +LLVGNK DILPKS+K  +
Sbjct: 57  DFLNILHGIGETDSLVVKIVDIPDFNGSWINGLQRLVGGNPILLVGNKADILPKSLKRER 116

Query: 121 VTQWLTERAHKEGLRPVDVILTSAQNHHAIKDLIDTIEKYRHGQDVYVVGVTNVGKSTLI 180
           + QW+   A E GL+PVDV L SA      I+++ID IE YR+G+DVYVVG TNVGKST I
Sbjct: 117 LIQWMKRKAKKLGLKPVDVFLVSAGRGQGIRELIDAIEHYRNGKDVYVVGCTNVGKSTFI 176

Query: 181 NAIIRELTGSRDVITISRFPGTTLDKIEIPLDDGSYIFDTPGIIHRHQMAHYLTAKNLKY 240
           N II+R++G  D+ITTS+FPGTTLD IEIPLDDGS ++DTPGII+ HQMAHY+  K+LK
Sbjct: 177 NRIIKEVSGKEDIITTSQFPGTTLDAIEIPLDDGSSLYDTPGIINNHQMAHYVNKDLKI 236

Query: 241 VSPKKEIKPKTYQLNSEQTLFLAGLARFDFISGQKQGFTAYFDNNLNLHRTKLVCADEFY 300
           +SPKKE+KP+T+QLN +QTL+  GLARFD++SG++   F  Y  N L +HRTKL  AD  Y
Sbjct: 237 LSPKKELKPRTFQLNDQQTLYPGGLARFDYVSGRRSPSTCYMPNELMIHRTKLRNADAIY 296

Query: 301 TKHVGKLLTPPTGKEVSDFPKLVRHRPTIKD-KMDIVYSGLGWIRVKSEAENPVVVAAWA 359
```

```
              KH G+LLTPP   E+ +FP+LV H FTIKD K DIV+SGLGW+ V    +    V A+A
Sbjct: 297 EKHAGELLTPPGKDEMDEFPELVAHTFTIKDKKTDIVPSGLGWVTVHDADKK---VTAYA 353

Query: 360 PEGVAVVLRKALI 372
           P+GV V +R++LI
Sbjct: 354 PKGVHVFVRRSLI 366
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5595> which encodes the amino acid sequence <SEQ ID 5596>. Analysis of this protein sequence reveals the following:

```
        Possible site: 15

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB14509 GB:Z99117 similar to hypothetical proteine [Bacillus subtilis]
 Identities = 220/373 (58%), Positives = 286/373 (75%), Gaps = 8/373 (2%)

Query: 1   MEELFCIGCGIQIQTEDKEKAGFTPAAALKKGMETGELYCQRCFRLRHYNEITDVHITDD 60
            ME++ CIGCG+ IQTEDK   G+ P A+L K      + CQRCFRL++YNEI DV +TDD
Sbjct: 1   MEKVVCIGCGVTIQTEDKTGLGYAPPASLTKE----NVICQRCFRLKNYNEIQDVSLTDD 56

Query: 61  EFLRLLHEVGDSDALVVNVIDIFDPNGSIIPGLSRFISGNDVLLVGNKKDILPKSVKDGK 120
           +FL +LH +G++D+LVV ++DIFDPNGS I GL R + GN +LLVGNK DILPKS+K  +
Sbjct: 57  DFLNILHGIGSTDSLVVKIVDIFDPNGSWINGLQRLVGGNPILLVGNKADILPKSLKRER 116

Query: 121 VTQWLTERAHEBGLRPDVMLTSAQNKYAIKDLIGRINELRNGRDVYVVGVTNVGKSTLI 180
           + QW+   A E GL+P+DV L SA     I+++I   RNG+DVYVVG TNVGKST I
Sbjct: 117 LIQWMKREAKELGLKPVDVFLVSAGRGQGIRRVIDAIEHYRNGKDVYVVGCTNVGKSTFI 176

Query: 181 NAIIQRITGNKDVITTSRFPGTTLDKIEIPLDDGTFIPDTPGIIHRHQMAHYLSPKELKI 240
           N II+E++G +D+ITTS+FPGTTLD IEIPLDDG+ ++DTPGII+ HQMAHY++ E+LKI
Sbjct: 177 NRIIKEVSGEEDIITTSQFPGTTLDAIEIPLDDGSSLYDTPGIINNHQMAHYVNKDLKI 236

Query: 241 VSPKKEIKPKTYQLNPEQTLFLGGLARFDFINGERQGFTAFFDNQLELHRTKLAGADAFY 300
           +SPKKE+KP+T+QLN +QTL+ GGLARFD+++GER  F  + N+L +HRTKL  ADA Y
Sbjct: 237 LSPKKELKPRTFQLNDQQTLYFGGLARFDYVSGERSPFICYMPNELMIHRTKLENADALY 296

Query: 301 DKHVGTLLTPPDKXRLTAFPKLVRHEPTI-DQKMDIVPSGLGWIRVNGQDSKAIVAAWA 359
           +KH G LLTPP K E+   FP+LV H FTI D+K DIVPSGLGW+ V+   D+   V A+A
Sbjct: 297 EKHAGELLTPPGKDEMDEFPELVAHTFTIKDKKTDIVPSGLGWVTVH---DADKKVTAYA 353

Query: 360 PEGVAVIVRKAII 372
           P+GV V VR+++I
Sbjct: 354 PKGVHVFVRRSLI 366
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 308/372 (82%), Positives = 343/372 (91%)

Query: 1   MEELFCIGCGARIQTENKDAAGYTPRAALEKGLETGELYCQRCFRLRHYNEITDVHITDD 60
           MEELFCIGCG +IQTE+K+ AG+TP AAL+KG+ETGELYCQRCFRLRHYNEITDVHITDD
Sbjct: 1   MEELFCIGCGIQIQTEDKEKAGFTPAAALKKGMETGELYCQRCFRLRHYNEITDVHITDD 60

Query: 61  EFLKLLHEVGDSDALVVNVIDIFDPNGSIIPGLSRPVAGNDVLLVGNKKDILPKSVKDGK 120
           EFL+LLHEVGDSDALVVNVIDIFDPNGSIIPGLsRP++GNDVLLVGNKKDILPKSVKDGK
Sbjct: 61  EFLRLLHEVGDSDALVVNVIDIFDPNGSIIPGLSRPISGNDVLLVGNKKDILPKSVKDGK 120

Query: 121 VTQWLTERAHEBGLRPDVILTSAQNHHAIKDLIDTIEKYRHGQDVYVVGVTNVGKSTLI 180
           VTQWLTERAHEBGLRP+DV+LTSAQN +AIKDLI  I + R+G+DVYVVGVTNVGKSTLI
Sbjct: 121 VTQWLTERAHEBGLRPLDVMLTSAQNKYAIKDLIGRINELRNGRDVYVVGVTNVGKSTLI 180
```

```
Query: 181 NAIIREITGSRDVITTSRFPGTTLDKIEIPLDDGSYIFDTPGIIHRHQMAHYLTAKNLKY 240
           NAII+EITG++DVITTSRFPGTTLDKIEIPLDDG++IFDTPGIIHRHQMAHYL+ K LK
Sbjct: 181 NAIIQEITGNKDVITTSRFPGTTLDKIEIPLDDGTPIFDTPGIIHRHQMAHYLSPKELKI 240

Query: 241 VSPKKEIKPKTYQLNSSQTLFLAGLARFDFISGQKQGFTAYFDNNLNLHRTKLVGADEFY 300
           VSPKKEIKPKTYQLN EQTLFL GLARFDFI+G++QGFTA+FDN L LHRTKL GAD FY
Sbjct: 241 VSPKKEIKPKTYQLNPEQTLFLGGLARFDFINGERQGFTAFFDNQLELHRTKLAGADAFY 300

Query: 301 TKHVGKLLTPPTGKEVSDFPKLVRHEPTIKDKMDIVYSGLGWIRVKSEAENPVVVAAWAP 360
            KHVG LLTPP  KE++ FPKLVRHEPTI  KMDIV+SGLGWIRV  + ++  +VAAWAP
Sbjct: 301 DKHVGTLLTPPDKKELTAFPKLVRHEPTIDQKMDIVPSGLGWIRVNGQKDSKAIVAAWAP 360

Query: 361 EGVAVVLRKALI 372
           EGVAV++KKA+I
Sbjct: 361 EGVAVIVRKAII 372
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1800

A DNA sequence (GBSx1907) was identified in *S.agalactiae* <SEQ ID 5597> which encodes the amino acid sequence <SEQ ID 5598>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2948(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14507 GB:Z99117 similar to dihydrodipicolinate reductase
          [Bacillus subtilis]
 Identities = 49/97 (50%), Positives = 67/97 (68%), Gaps = 2/97 (2%)

Query: 1  MLTSKQRAFLKSEAHSMKPIIQIGKNGLNDQIKTSVKNALDARELIKVTLLQNTDEDIHD 60
           MLT KQ+ FL+S+AH + PI Q+GK G+ND +   + AL+ARELIKV++LQN +ED +D
Sbjct: 1  MLTGKQKRFLRSKAHHLTPIPQVGKGGVNDNMIKQIAEALEARKLIKVSVLQNCREDKND 60

Query: 61 VAEVLEDEIGCTVTVLKIGRILLLYKKSARKENRKISV 97
           VAE L      V  IG  ++LYKES  KEN++I +
Sbjct: 61 VAEALVKGSRSQLVCTIGNTIVLYKES--KENKQIEL 95
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5599> which encodes the amino acid sequence <SEQ ID 5600>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2839(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 89/102 (87%), Positives = 98/102 (95%)

Query: 1   MLTSKQRAFLKSEAHSMKPIIQIGKNGLNDQIKTSVRNALDARELIKVTLLQNTDEDIHD 60
            MLTSKQRAFLKSEAHS+KPI+QIGKNGLND IKTS+R ALDARELIKVTLLQNTDEDIH+
Sbjct: 1   MLTSKQRAFLKSEAHSLKPIVQIGKNGLNDHIKTSIRQALDARELIKVTLLQNTDEDIHE 60
```

```
Query: 51  VAEVLEDEIGCDTVLKIGRILILYKESAKKENRKISVKVKAV 102
           VAE+LE+EIGCDTVLKIGRILILYK SA+KENRK+S KVKA+
Sbjct: 61  VAEILEEEIGCDTVLKIGRILILYKVSAKKENRKLSPKVKAI 102
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1801

A DNA sequence (GBSx1908) was identified in *S.agalactiae* <SEQ ID 5601> which encodes the amino acid sequence <SEQ ID 5602>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.66   Transmembrane    3 -  19 (   1 -  21)

----- Final Results -----
                bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10089> which encodes amino acid sequence <SEQ ID 10090> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14506 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 85/187 (45%), Positives = 134/187 (71%)

Query: 38  KQIGIDGGNFNFVFNAHLVVADQVRQQLCLDQVLLMPEFQPPHIDKKETIDEQHRLKMLE 97
           K+IGI GG F+F  N HL++A++V  Q  LD++  MP   PPH    ++  D  HR++ML+
Sbjct: 2   KKIGIFGGTFDPFHNGHLLMANEVLYQAGLDEINFMPNQIPPHKQNEDYTDSFHRVEMLK 61

Query: 98  LAIEGIDGLSIEPIEIERKGISYTYDTMKLLIEKNPDVDYYFIIGADMVEYLPKWHRIDE 157
           LAI+       +E +E+ER+G SYT+DT+ LL ++ P+    +FIIGADM+EYLPKW++++DE
Sbjct: 62  LAIQSNPSFKLELVEMEREGPSYTFDTVSLLKQRYPNDQLFFIIGADMIEYLPKWYKLDE 121

Query: 158 LVKMVQFVGVQRPKYKAGTSYPVIWVDLPLMDISSSMIRQFIKSNRQPNYLLPREVLDYI 217
           L+ ++QF+GV+RP +   T YP++ D+P  ++SS+MIR+  KS +  +YL+P +V  Y+
Sbjct: 122 LLNLIQFIGVKRPGPHVETPYPLLFADVPEPEVSSTMIRERFKSKKPTDYLIPDKVKKYV 181

Query: 218 RKEGLYK 224
           + GLY+
Sbjct: 182 EENGLYE 188
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5603> which encodes the amino acid sequence <SEQ ID 5604>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.4660(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 172/210 (81%), Positives = 196/210 (92%)

Query: 15  MALELLTPFTKVELEEKKRDTNRKQIGIMGGNFNPVHNAHLVVADQVRQQLCLDQVLLMP 74
           MALELLTPFTKVELEE+K+++NRKQIGI+GGNFNP+HNAHLVVADQVRQQL LDQVLLMP
Sbjct: 1   MALELLTPFTKVELEEEKKRSNRKQIGILGGNFNPIHNAHLVVADQVRQQLGLDQVLLMP 60
```

```
Query: 75  EFQPPHIDKKETIDEQHRLKMLELAIBGIDGLSIEPIEIERKGISYTYDTMKLLIEKNPD 134
           E +PPH+D KETIDE+HRL+MLELAIE ++GL+IE  E+ER+GISYTYDTM  L E++PD
Sbjct: 61  ECKPPHVDAKETIDEKHRLRMLELAIEDVEGLAIETCKLERQGISYTYDTMLYLTEQHPD 120

Query: 135 VDYYFIIGADMVEYLPKWHRIDELVKMVQFVGVQRPKYKAGTSYPVIWVDLPLMDISSSM 194
           VD+YFIIGADMV+YLPKWHRIDELVK+VQFVGVQRPKYKAGTSYPVIWVDLPLI+ISSSM
Sbjct: 121 VDFYFIIGADMVDYLPKWHRIDELVKLVQFVGVQRPKYKAGTSYPVIWVDLPLIDISSSM 180

Query: 195 IRQFIKSNRQPNYLLPREVLDYIRKEGLYK 224
           IR FIK  RQPNYLLP+ VLDYI +EGLY+
Sbjct: 181 IRDFIKKGRQPNYLLPKRVLDYITQEGLYQ 210
```

SEQ ID 5602 (GBS651) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 8-10; MW 53.3kDa) and in Figure 186 (lane 8; MW 53kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 12; MW 28.4kDa) and in Figure 140 (lane 11; MW 20kDa).

Purified GBS651-GST is shown in Figure 243, lane 4; purified GBS651-His is shown in Fig.229, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1802

A DNA sequence (GBSx1909) was identified in *S.agalactiae* <SEQ ID 5605> which encodes the amino acid sequence <SEQ ID 5606>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4281(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14505 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 79/180 (43%), Positives = 115/180 (63%)

Query: 9   LDRTELLSKVRHMMSDKRFNHVLGVERAAIELAERYGYDKEKAGLAALLHDYAKELSDDE 68
            ++R E L+ V+  +++ R+ H +GV   AIELAER+G D +KA +AA+ HDYAK    +E
Sbjct: 1   MNREEALACVKQQLTEHRYIHTVGVMNTAIELAERFGADSKKAEIAAIFHDYAKFRPKEE 60

Query: 69  FLRLIDKYQPDPDLKKWGNNIWHGLVGIYKIQEDLAIKDQDILAATAKHTVGSAQMSTLD 128
              ++I + +     L       +WH   VG Y +Q +   ++D+DIL AI  HT G   M+ L+
Sbjct: 61  MKQIIAREKMPAHLLDHNPELWHAPVGAYLVQREAGVQDEDILDAIRYHTSGRPGMTLLE 120

Query: 129 KIVYVADYIEHNRQFPGVEEARKLAKVDLNKAVAYETARTVAFLASKAQPIYPKTIETYN 188
           K++YVADYIE NR FPGV+E R+LA+ DLN+A+      T+ FL  K QP++P T  TYN
Sbjct: 121 KVIYVADYIEPNRAFPGVDEVRKLARTDLNQALIQSIKWTMVFLMKKNQPVFPDTFLTYN 180
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5607> which encodes the amino acid sequence <SEQ ID 5608>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2615(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 130/194 (67%), Positives = 159/194 (81%)

Query: 1   MTYKDYTGLDRTELLSKVRHMMSDKRPNHVLGVERAAIELAERYGYDKEKAGLAALLHDY 60
            MTY+DY    RTELL+K+   MS KRP HVLGVE+AA+ LAE YG + +KAGLAALLHDY
Sbjct: 1   MTYEDYLPYSRTELLAKIAEQMSPKRPKHVLGVEKAALSLAECYGCNPDKAGLAALLHDY 60

Query: 61  AKELSDDEFLRLIDKYQPDPDLKKWGNNIWHGLVGIYKIQRDLAIKDQDILAAIAKHTVG 120
           AKS  D  FL LIDKYQ  P+L KW NN+WHG+VGIYKIQRDL +KD+DIL AI  HTVG
Sbjct: 61  AKSCPDQVFLDLIDKYQLSPELAKWNNNVWHGMVGIYKIQRDLGLKDKDILRAIEIHTVG 120

Query: 121 SAQMSTLDKIVYVADYIEHNRDFPGVSEARELAKVDLNKAVAYETARTVAFLASKAQPIY 180
           +A+M+ LDK++YVADYIE  R FP V++AR++AK+DLN+AVAYET  TVA+LASKAQPI+
Sbjct: 121 AAEMTLLDEVLYVADYIEEGRIFPLVDDARKIAKLDLNQAVAYETVNTVAYLASKAQPIF 180

Query: 181 PKTIETYNAYIPYL 194
            P+T++TYNA+  YL
Sbjct: 181 PQTLDTYNAFCSYL 194
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1803

A DNA sequence (GBSx1910) was identified in *S.agalactiae* <SEQ ID 5609> which encodes the amino acid sequence <SEQ ID 5610>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.34   Transmembrane   12 -  28 (  10 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10087> which encodes amino acid sequence <SEQ ID 10088> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG19496 GB:AE005041 Vng1100c [Halobacterium sp. NRC-1]
 Identities = 46/175 (26%), Positives = 82/175 (46%), Gaps = 12/175 (6%)

Query: 22  ALLLIDIQQGIMDKK--PKHLTNPAVLLDDLLLSAKGSNCEVIWIRHHDKS----LPQGS 75
           AL+L+D QQG D      ++  +     ++LL + + +   +RH+  E    L QG
Sbjct: 7   ALVLVDFQQGFADPAWGDRNNPDASAHAEELLAAWRDAAAAPIAHVRHNSTEATSPLRQGE 66

Query: 76  PQWEIWEQRHLVTHHKIIDKTYNSCFKDTHLHDYLQSKHISQLIMMGLQTEYCVDTSVKV 135
           P +   +              K+ N  F DT L  +L+ +     L++ GL T++C  T+V++
Sbjct: 67  PGFAYTDGLAPAADEPEPVKSVNGAFVDTALEGWLRDRDIGSLVVCGLTTDHCVSTTVRM 126

Query: 136 AFEYGYDIFIPQGGHLTFDTPTLSGDSIKK---HYENIWHHR--FATMVAKDSLL 185
           A  G+D+ + +    T D  TL G+ +     H   + H R  FAT+     ++L
Sbjct: 127 ADNRGFDVTLVRDATATHDR-TLDGERLPSVVHRTALAHLRGEFATLATTATVL 180
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5610 (GBS652) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 133 (lane 2 & 3; MW 49.7kDa) + lane 4; MW 27kDa) and in Figure 186 (lane 9; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell

extract is shown in Figure 133 (lane 5 & 7; MW 24.8kDa) and in Figure 178 (lane 10; MW 25kDa). Purified GBS652-GST is shown in Figure 243, lane 9; purified GBS652-His is shown in Figure 229, lane 10.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1804

A DNA sequence (GBSx1911) was identified in *S.agalactiae* <SEQ ID 5611> which encodes the amino acid sequence <SEQ ID 5612>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0945(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14504 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 55/118 (46%), Positives = 82/118 (68%)

Query: 1   MTEKDLLQLVVKAADEKRAEDIVILDLQFVTSVADYFVIMSASNSRQLEAIADNIREQVK 60
            M +K +L++    A D+KRADDI+ LD++ ++ VADYF+I    ++ +Q++AIA  I++Q
Sbjct: 1   MNQKSILKIAAAACDDKRADDILALDMEGISLVADYFLICHGNSDKQVQAIAREIKDQAD 60

Query: 61  GNGGDASHLEGDSKAGWVLLDLNSVVVHIFSEDERQHYNLEKLWHEAPLLDAEVFMTE 118
             NG      +EG  +A WVL+DL  VVVH+F +DER +YNLEKLW +APL D +  M +
Sbjct: 61  ENGIQVKKMEGFDEARWVLVDLGDVVVHVFHKDERSYYNLEKLWGDAPLADLDFGMNQ 118
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5613> which encodes the amino acid sequence <SEQ ID 5614>. Analysis of this protein sequence reveals the following:

```
      Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.69   Transmembrane   91 - 107 (  91 - 107)

----- Final Results -----
             bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB14504 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 55/113 (48%), Positives = 80/113 (70%)

Query: 17  MKKEELLKIVVEATKEKRADILALDLEGLTSLVDYFVIASATNSRQLEAIADNIREKVK 76
            M ++ +LKI    A ++KRA+DILALD+EG++ + DYF+I    + +Q++AIA  I+++
Sbjct: 1   MNQKSILKIAAAACDDKRADDILALDMEGISLVADYFLICHGNSDKQVQAIAREIKDQAD 60

Query: 77  EAGGDASHVEGNSQAGWVLLDLTDVVVHLFLEDERYHYNLEKLWHEAPAVALD 129
            E G      +EG  +A WVL+DL DVVVH+F +DER +YNLEKLW +AP    LD
Sbjct: 61  ENGIQVKKMEGFDEARWVLVDLGDVVVHVFHKDERSYYNLEKLWGDAPLADLD 113
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 78/116 (67%), Positives = 100/116 (85%)

Query: 1   MTEKDLLQLVVKAADEKRAEDIVILDLQFVTSVADYFVIMSASNSRQLEAIADNIREQVK 60
```

```
            M +++LL++VV+A +EKRA+DI+ LDL+ +TS+ DYPVI SA+NSRQLEAIADNIRR+VK
Sbjct: 17  MKKEELLKIVVEATEEKRAKDILALDLEGLTSLTDYPVIASATNSRQLEAIADNIREKVK 76

Query: 61  GNGGDASHLEGDSKAGWVLLDLNSVVVHIPSEDERQHYNLEKLWHEAPLLDAEVFM 116
            GGDASH+EG+S+AGWVLLDL  VVVH+P EDER HYNLEKLWHEAP +  + ++
Sbjct: 77  EAGGDASHVSGNSQAGWVLLDLTDVVVHLPLEDERYHYNLEKLWHEAPAVALDAYL 132
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1805

A DNA sequence (GBSx1912) was identified in *S.agalactiae* <SEQ ID 5615> which encodes the amino acid sequence <SEQ ID 5616>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2415(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1806

A DNA sequence (GBSx1913) was identified in *S.agalactiae* <SEQ ID 5617> which encodes the amino acid sequence <SEQ ID 5618>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1570(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14503 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 86/242 (35%), Positives = 154/242 (63%), Gaps = 4/242 (1%)

Query: 4   YETFAAVYDAVMDDTLYAKWTDFSLRHFPKGKKKLLELACGTGIQSVRFAQAGYAVTGLD 63
            Y+ FA+VYD +M     Y +WT +       P+ K ++L+LACGTG  S+R A+ G+ VTG+D
Sbjct: 3   YQGFASVYDELMSHAPYDQWTKWIEASLPE-KGKILDLACGTGEISIRLAEKGFEVTGID 61

Query: 64  LSGDMLKLAKKRATSAHQSIQFIEGNMLDLSNV-GKYDLITCYSDSIGYMQDEVEGDVF 122
            LS +ML  A+++ +S+ Q I F++ +M +++    G++D +    DS+ Y++ + +V + F
Sbjct: 62  LSEEMLSFAQQKVSSS-QPILFLQQDMREITGFIGQPDAVICCDSLNYLKTKNDVIETF 120

Query: 123 IEVYKALEENGVFIFDVHSTYQTDKVFPGYSYHENADDFAMVWDTYEDDAPHSIVHELTF 182
              V++ L+  G+ +FDVHS+++  +VFP  ++ +  +D + +W ++       S++H+++F
Sbjct: 121 KSVPRVLKPEGILLFDVHSSFKIAEVFPDSTFADQDEDISYIWQSFAGSDELSVIHDMSF 180

Query: 183 FVQEEDGRFTRHDEVHEERTYDILTYDILLELQAGFKDVKVYADFEDKKPTATSARWFFVA 242
            FV   + + R DE HE+RT+ +  Y+ +L+  GF+  +V ADF D +P+A S R FF A
Sbjct: 181 FVWNGEA-YDRFDETHEQRTFFVSEVEEMLKNCGFQLHRVFADFTDTEPSAQSERLFFKA 239
```

```
Query: 243 HK 244
           K
Sbjct: 240 QK 241
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5619> which encodes the amino acid sequence <SEQ ID 5620>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2315(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 191/243 (78%), Positives = 215/243 (87%), Gaps = 2/243 (0%)

Query: 4   YETPAAVYDAVMDDTLYAKWTDFSLRHFPK--GKKKLLELACGTGIQSVRFAQAGVAVTG 61
           YE FA+VYDAVMDD+LY  WTDFSLRH PK  G+  +LLELACGTGIQSVRFAQAG+ VTG
Sbjct: 21  YSKFASVYDAVMDDSLYDLWTDFSLRHLPKSRGRNRLLELACGTGIQSVRFAQAGFDVTG 80

Query: 62  LDLSGDMLKLAKKRATSAHQSIQFIEGNMLDLSNVGKYDLITCYSDSICYMQDEVEVGDV 121
           LDLS  ML +AKKRA SA + I FI+GNMLDLS VG++D +TCYSDSICYMQDEV+VGDV
Sbjct: 81  LDLSQDMLAIAKKRAQSAKKKIDFIQGNMLDLSQVGQFDEVTCYSDSICYMQDEVDVGDV 140

Query: 122 FIEVYKALEENGVFIFDVHSTYQTDKVFPGYSYHENADDFAMVWDTYEDDAPHSIVHELT 181
           F EVY  L  +G+FIFDVHSTYQTD+ FPGYSYHENADDFAMVWDTY D+APHS+VHELT
Sbjct: 141 FKEVYDVLANDGIFIFDVHSTYQTDECFPGYSYHENADDFAMVWDTYADEAPHSVVHELT 200

Query: 182 FFVQEEDGRFTRHDEVHEERTYDILTYDILLTQAGFKDVKVYADFEDKKPTATSARWFFV 241
           FF+QE+DGRF+R DEVHEERTY++LTYDILLEQAGFK  KVYADFEDK+PT TS RWFFV
Sbjct: 201 FFIQEDDGRFSRFDEVHEERTYELLTYDILLEQAGFKSPKVYADFEDKEPTKTSKRWFFV 260

Query: 242 AHK 244
           A+K
Sbjct: 261 AYK 263
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1807

A DNA sequence (GBSx1914) was identified in *S.agalactiae* <SEQ ID 5621> which encodes the amino acid sequence <SEQ ID 5622>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3538(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06304 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 129/367 (35%), Positives = 184/367 (49%), Gaps = 45/367 (12%)

Query: 1   MTVTGIVAEPNPFHNGHKYLLEQAQ-----GIKVIAMSGNFMQRGEPAIVDKWTRSQMAL 55
           M   G+V E+NPFHNGH + L +A+       + +  MSG F+QRGEPAI+ KW R+ +AL
Sbjct: 1   MKAVGVVVEYNPFHNGHLHHLITEARKQAKADVVIAVMSGYFLQRGEPAILPKWERTSLAL 60
```

```
Query: 56  ENGADLVIELPPLVSVQSADYFASGAVSILARLGVDNLCPGTEE--MLDYARIGDIYVNK 113
           + GADLV+ELP+  S Q A++FA+GAVSILA L  D LCPG+EE  +  + R+
Sbjct: 61  QGGADLVVELPYAPSTQKAEWFATGAVSILAALEADALCPGSERGTIEPPHRLYFMAKH 120

Query: 114 KEEMEAFLKKQSD-SLSYPQKMQAMWQEFAGIT--FSGQTPNHILGLAYTKAA--SQNGI 168
           +   +  +K++ D  +SYP       ++   G          PN+ILG   Y KA         I
Sbjct: 121 RLAWDRMIKEELDKGMSYPTATSLAFKRLEGSAEHLDLSRPNNILGFHYVKAYDLHTSI 180

Query: 169 RLNFIQRQGAGYHSSEKTE-IFASATSLRK--------HQSDRFF------VEKGMPNSD 213
           +   I R  AGYH     E   ASATS+RK           DR            + K
Sbjct: 181 KAMTIPRIKAGYHDDSLNESSIASATSIRKSLKTKEGWQMVDRVVPSYTTEMLKSFEKET 240

Query: 214 LFLNSPQVVWQDYFSLLKYQIMTHS--DLTQIYQVNEEIANRIKSQIRYVETVDKLVDKV 271
             FL S    W+  F LLKY++T +   L  IY+  E +  R    I    +  + K+
Sbjct: 241 TFLPS----WERLFPLLKYRLLTATPEQLHAIYSGEEGLEYRALKTTVSATSPHDWMTKM 296

Query: 272 ATKRYTKARIRRLLTYILINAVESPIPNA----------IHVLGFTQKGQQHLKSVKK-- 319
             TKRYT  RI+R  T++  N  +  I +           I +LG T +GQ +L   KK
Sbjct: 297 KTKRYTWTRIQRYATHLFTNTTKEIHSVLPRGTESLPYIRLLGMTSRGQMYLNGKKKQL 356

Query: 320 SVDIVTR 326
           +  ++TR
Sbjct: 357 TTPVITR 363
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5623> which encodes the amino acid sequence <SEQ ID 5624>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3165(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 221/359 (61%), Positives = 288/359 (79%)

Query: 1   MTVTGIVAEFNPFHNGHKYLLEQAQGIKVIAMSGNFMQRGEPAIVDKWTRSQMALENGAD 60
           MTVTGI+AEFNPFHNGHKYLLE A+G+K+IAMSGNFMQRGEPA++DKW RS+MAL+NGAD
Sbjct: 1   MTVTGIIAEFNPFHNGHKYLLETAEGLKIIAMSGNFMQRGEPALIDKWIRSEMALKNGAD 60

Query: 61  LVIELPLVSVQSADYFASGAVSILARLGVDNLCPGTEEMLDYARIGDIYVNKKEEMEAF 120
           +V+ELPF VSVQSADYFA GA+ IL +LG+  L PGTE ++DY ++  +Y  K E+M A+
Sbjct: 61  IVVELPFFVSVQSADYFAQGAIDILCQLGIQQLAFGTENVIDYQKLIKYVEKKSEQMTAY 120

Query: 121 LKKQSDSLSYPQKMQAMWQEFAGITFSGQTPNHILGLAYTKAASQNGIRLNFIQRQGAGY 180
           L    D+ SYPQK Q MW+ FAG+ FSGQTPNHILGL+Y KA++    I+L PI+RQGA Y
Sbjct: 121 LSTLEDTFSYPQKTQKMWEIFAGVKFSGQTPNHILGLSYAKASAGKHIQLCPIKRQGAAY 180

Query: 181 HSSEKTEIFASATSLRKHQSDRFFVEKGMPNSDLFLNSPQVVWQDYFSLLKYQIMTHSDL 240
           HS +K  + ASA+++R+H +D  F+    +PN+ L +N+P + W  YFS LKYQI+ HSDL
Sbjct: 181 HSKDKNHLIASASAIRQHLNDWDFISHSVPNAGLLINNPHMSWDHYFSFLKYQILNHSDL 240

Query: 241 TQIYQVNEEIANRIKSQIRYVETVDELVDKVATKRYTKARIRRLLTYILINAVESPIPNA 300
           T I+QVN+E+A+RIK  I+  + +D LVD VATKRYTKAR+RR+LTYIL+NA E  +P
Sbjct: 241 TSIFQVNDELASRIKKAIKVSQNIDHLVDTVATKRYTKARVRRILTYILVNAKEPTLPKG 300

Query: 301 IHVLGFTQKGQQHLKSVKKSVDIVTRIGSQTWDSLTQRADSVYQMGNANIARQTWGRIP 359
           IH+LGFT KGQ HLK +KKS  ++TRIG++TWD +TQ+ADS+YQ+G+ +I EQ++GRIP
Sbjct: 301 IHILGFTSKGQAHLKKLKKSRPLITRIGARTWDEMTQKADSIYQLGHQDIPEQSPGRIP 359
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1808

A DNA sequence (GBSx1915) was identified in *S.agalactiae* <SEQ ID 5625> which encodes the amino acid sequence <SEQ ID 5626>. This protein is predicted to be transcriptional activator tipa. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3117(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15677 GB:Z99122 transcriptional regulator [Bacillus subtilis]
 Identities = 91/246 (36%), Positives = 144/246 (57%), Gaps = 14/246 (5%)

Query: 4   VKEISHISGISVRTLHYYDEIDLLSPSFVGENGYRYYDDESLIKLQEILLFKELEFPLKK 63
           VK+++ ISG+S+RTLH+YD I+LL+PS + + GYR Y D  L +LQ+IL FKE+ F L +
Sbjct: 5   VKQVAEISGVSIRTLHHYDNIELLNPSALTDAGYRLYSDADLERLQQILFFKKIGFRLDE 64

Query: 64  IKEIMDSPNYDRNQALLDQIRWLELKKQRLEEVIEHAK----SIQRGKNMSD---FTAYN 116
           IKE++D PN+DR  AL  Q   L  KKQR++E+I+       S+  G+ M+     F   +
Sbjct: 65  IKEMLDHPYFDRKAALQSQKSIIAWKKKQRMDEMIQTIDRTLLSVDGGETMNKRDLFAGLS 124

Query: 117 QKELEAFQ----EEARTRWGD--TDSYKEPENSHSKNDFSMISQAMSQIFKDFGQLKELS 170
            +++E  Q    +E R  +G    +  ++  +++S +D+  I      I++       +
Sbjct: 125 MKDIEEHQQTYADEVRKLYGKEIAEETEKRTSAYSADDWRTIMAEFDSIYRRIAARMKKHG 184

Query: 171 PTDEKVQKQVQILQDYITAQFYNCTNDLLASLGIMYIQDERFQKSIDNWGGQGTALFVSK 230
           P D ++Q  V   +D+I    Y+CT D+   LG +YI DERF  SI+ + G+G A F+ +
Sbjct: 185 PDDAEIQAAVGAFRDHICQYHYDCTLDIPRGLGEVYITDERFTDSINQY-GEGLAAFLRE 243

Query: 231 AIDSYC 236
           AI  YC
Sbjct: 244 AIIIYC 249
```

There is also homology to SEQ ID 1712.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1809

A DNA sequence (GBSx1916) was identified in *S.agalactiae* <SEQ ID 5627> which encodes the amino acid sequence <SEQ ID 5628>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14597 GB:Z99117 yrkC [Bacillus subtilis]
 Identities = 56/129 (43%), Positives = 74/129 (56%), Gaps = 7/129 (5%)

Query: 2   KGFHGNIEKKLTLGNTNFRQVLYTAEHCQLVLMTLPVGGEIGSEIHAENDQFFRFEAGHGK 61
           K F  NI + T  N  FR  L+T +H C+ LM+L +G+IG EIH   DQF R E G G
```

```
Sbjct: 59  KPFVVNINRATKQNNTFRTALWTGKHPQVTLMSLGIGRDIGLEIHPNVDQFLRIRQGRGI 118

Query: 62  VVIDGN------RYEVADGDAIIVPAGAEHNVINTSETEMLKLYTIYSPAHHKDGIRAT 115
           V +  +      + V D  AI+VPAG  HNVINT   T  LKLY+IY+P +H  G +  T
Sbjct: 119 VKMGKSKDHLNFQRNVVDDSAIVVPAGTWHNVINTGNTP-LKLYSIYAPPNHPFGTVHET 177

Query: 116 RKEAEENER 124
           + +A   E+
Sbjct: 178 KADAVAAED 186
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1810

A DNA sequence (GBSx1917) was identified in *S.agalactiae* <SEQ ID 5629> which encodes the amino acid sequence <SEQ ID 5630>. This protein is predicted to be glycerol uptake facilitator (glpF). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.08   Transmembrane  156 - 172 ( 153 - 180)
   INTEGRAL    Likelihood = -6.21   Transmembrane  135 - 151 ( 132 - 155)
   INTEGRAL    Likelihood = -4.09   Transmembrane   86 - 102 (  80 - 103)
   INTEGRAL    Likelihood = -3.93   Transmembrane  213 - 229 ( 212 - 230)
   INTEGRAL    Likelihood = -3.72   Transmembrane    8 -  24 (   5 -  28)
   INTEGRAL    Likelihood = -2.76   Transmembrane   38 -  54 (  36 -  58)

----- Final Results -----
               bacterial membrane --- Certainty=0.4630(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04811 GB:AP001510 glycerol uptake facilitator [Bacillus halodurans]
 Identities = 135/230 (58%), Positives = 171/230 (73%)

Query: 1   MTQFLGEFLGTFILVLLGDGVVAGNVLSKTKEEGTGWTAIVFGWGIACTVAVYVSGLFSP 60
           M+ FLGE +GT IL++LG GVVAG VL  TK E  GW  I   WG+A   AVY  G  S
Sbjct: 1   MSPFLGEVIGTMILIILGGGVVAGVVLKGTKSENGGWIVITAAWGLAVATAVYCVGQISG 60

Query: 61  AHLNPAVTLAWASIGAISWGQVIPYIAQMLGAMVAATTLWLHYYPHWKETKDSGLILAS 120
           AHLNPAVT+ +A +GA  W QV  +I+AQMLGAM+ AT+++LHYYPH+K T+D G  LA
Sbjct: 61  AHLNPAVTIGLALVGAFEWSQVAGYIVAQMLGAMIGATLVFLHYYPHFKATEDQGAKLAV 120

Query: 121 FSTGPAIRHTPSNLLGEIIGTAILVITIMAIGPSKVAAGLGPIIVGIVIFAVGFSLDPTT 180
           FST PAI+H P+N   E++GT +LV+ I+AIG ++   GL P+IVG++I  +G SL  TT
Sbjct: 121 FSTDPAIKHLPANFFSEVLGTFVLVLGILAIGANEFTEGLNPLIVGLLIVVIGLSLGGTT 180

Query: 181 GYAINPARDLGPRLMHAILPIENKGNSDWGYAWIPVVGPIIGGVLGAILY 230
           GYAINPARDLGPR+ H +LPI  KG+S+WSYAWIP+VGPIIGG +GA+ Y
Sbjct: 181 GYAINPARDLGPRIAHFLLPIPGKGSSNWSYAWIPIVGPIIGGGIGALTY 230
```

There is also homology to SEQ ID 2854.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1811

A DNA sequence (GBSx1918) was identified in *S.agalactiae* <SEQ ID 5631> which encodes the amino acid sequence <SEQ ID 5632>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1694(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07114 GB:AP001518 unknown conserved protein in others
            [Bacillus halodurans]
 Identities = 64/118 (54%), Positives = 85/118 (71%)

Query: 5   GIIVVSHSKNIAQGVVDLISEVAKDVSITYVGTEDGEIGTSYDQVQQIVEQNDKKTLLA 64
            GI++ SH   +A+G+V L+ E AKDVSITY GGT+D ++G SP+++QQ V  N+   L
Sbjct: 7   GIVISSHVPALAEGIVTLLKEAAKDVSITYAGGTDDDQWGASPEKIQQAVMDNEADELFV 66

Query: 65  FFDLGSAKMNLELVADFSEKNIIINSVPVVEGAYTAAALLQAGADLDSIQSQLAELTI 122
            F+DLGSAKM+E+V + SEK I +  V +VEGAYTAAAL Q GA  ++I  QL  LTI
Sbjct: 67  FYDLGSAKMVEMVMELSEKTIHLMDVALVEGAYTAAALTQGGASFETIMEQLQPLTI 124
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1812

A DNA sequence (GBSx1919) was identified in *S.agalactiae* <SEQ ID 5633> which encodes the amino acid sequence <SEQ ID 5634>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4753(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07115 GB:AP001518 unknown conserved protein in others
            [Bacillus halodurans]
 Identities = 98/190 (51%), Positives = 135/190 (70%), Gaps = 2/190 (1%)

Query: 3   VKTAIEWMHTFNQKIQSNKDYLSKLDTPIGDGDHGSNMARGMTAVIENLDNNEFSSAADV 62
            V+   +W+H F++K+Q+N+ YLSELD+ IGDGDHG NMARG+  V   L  N F S  +V
Sbjct: 4   VENTTKWLHAFHEKVQANQSYLSSLDSAIGDGDHGTNMARGLAEVERKLKENLFBSPQEV 63

Query: 63  FKTVSMQLLSKVGGASGPLYGSAPMGITK-AEQSKSTISEALGALEMIQKRGKAELNEK 121
             K  +M L+SK GGAGGPLYG+A  + ++K          I +++ AGL  I KRGKA    EK
Sbjct: 64  LKMAAMALISKTGGASGPLYGTALLEMSKQVANDPQNIGKSTEAGLNGILKRGKATTGEK 123

Query: 122 TMVDVWHGVIEAI-EKNELMTEDRIDSLVDATKGMKATKGRASYVGERSVGHIDPGSFSSG 180
            TKVD+W  V+E++  + +L+++RI   V  TK MKATKGRASY+GERS+GH+DPG+ SSG
Sbjct: 124 TMVDIWKPVVESLMAEQQLSKERIQQPVSETKEMKATKGRASYLGERSLGHLDPGAVSSG 183

Query: 181 LLFKALLEVG 190
             LF+A+++ G
```

```
Sbjct: 184 YLFEAMIDGG 193
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1813

A DNA sequence (GBSx1920) was identified in *S.agalactiae* <SEQ ID 5637> which encodes the amino acid sequence <SEQ ID 5638>. This protein is predicted to be dihydroxyacetone kinase (b1200). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2080(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07116 GB:AP001518 dihydroxyacetone kinase [Bacillus halodurans]
 Identities = 204/329 (62%), Positives = 261/329 (79%)

Query: 1   MKKILNQPTDVVTEMLDGLAYVHNDLVHRIEGFDIIARNEEKSGKVALISGGGSGHEPSH 60
           MKKILN P +V+ EMLDG  Y +  LV R+ G  +I R  E  GKVAL+SGGGSGHEPSH
Sbjct: 1   MKKILNDPQNVLDEMLDGFVYANGHLVERVAGTGVIRRTYEDKGKVALVSGGGSGHEPSH 60

Query: 61  AGFVGEGMLSAAVCGAVFTSPTPDQVLKAIKEADEGAGVFMVIKNYSGDIMNFEMAQDMA 120
           AGFVG+GMLSAAVCG VPTSPTPDQ+ E IK AD+G GV ++IKNY+GD+MNFEMA +MA
Sbjct: 61  AGFVGQGMLSAAVCGEVPTSPTPDQIFEGIKAADQGGGVLLIIKNYTGDVMNFEMAGEMA 120

Query: 121 EMBGIKVASVVVDDDIAVEDSLYTQGKRGVAGTILVHKILGHAARHGKSLQEIKAIADEL 180
           E EGI V  ++V+DDIAVEDS +T G+RGVAGTI+VHKI+G AA  G SLQ +K + + +
Sbjct: 121 EABGITVDHIVNDDIAVEDSSFTAGRRGVAGTIIVHKIVGAAAEAGLSLQSLKVLGETV 180

Query: 181 VPNIHTVGLALSGATVPEVGKPGFVLAEDEIEFGIGIHGEPGYRKEKMQPSKALATELVD 240
           + N  T+G+++  ATVP VGKPGF L +DE+E+G+GIHGEPGYRKEK++ SK +A EL+
Sbjct: 181 IENTKTIGVSILPATVPAVGKPGFSLGDDENEYGVGIHGEPGYRKEKLKSSKEIASELIL 240

Query: 241 KLIESFDAKSGHKYGVLINGMGATPLMEQYVFANDVAKLLEDKGIEVNYKKLGNVNTSID 300
           KL E+F     G+KYGVL+NG+GATPLMEQYVF NDVA  L ++G+ + +KK+G++NTSID
Sbjct: 241 KLKEAFGWSKGDKYGVLVNGLGATPLMEQYVFWNDVANKLTEEGLNIQFKKVGSFMTSID 300

Query: 301 MAGLSLTLIKLENQEWLEALNSDVTTIAW 329
           MAG+SLTLIK+  ++WL+  N +V T+ W
Sbjct: 301 MAGVSLTLIKIVEEKWLDYWNHSVKTVDW 329
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1814

A DNA sequence (GBSx1921) was identified in *S.agalactiae* <SEQ ID 5639> which encodes the amino acid sequence <SEQ ID 5640>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1997(Affirmative) < succ>
```

```
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07113 GB:AP001518 unknown [Bacillus halodurans]
 Identities = 59/142 (41%), Positives = 82/142 (57%), Gaps = 5/142 (3%)

Query: 1   MTSSLITKKKIAKSFKRLFISQAFDKISVSDIMEDAGIRRQTFYNHFVDKYALLEWIFQT 60
            MT+S+ITKK IAK+FK L   Q F KISVSDIM  A +RKQTFY HF DK+ LL WI++
Sbjct: 1   MTNSIITKKVIAKAFKDLMEVQFPSKISVSDIMNRANMRKQTFYTHFQDKFELLEWIYKQ 60

Query: 61  ELSEQVTDNLDYISGFQLLSELLTFFKMNQSFYIKLFQIEDQNDFSSYFESYCEQLVDKL 120
            E  E   D L Y     +  L+ +F  NQ FY +   + QN F+ Y   + + L
Sbjct: 61  ETKEHSIDFLAYDDLHTIFRHLMHYFVENQTFYQRAWVNGQMGFTDYLYEHIQTL---Y 117

Query: 121 LSDYSKSNFWQKERVTFINYHS 142
            L++  +  +QK+R    +++S
Sbjct: 118 LNEIDRR--SQKDREFISSFYS 137
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5641> which encodes the amino acid sequence <SEQ ID 5642>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2101(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 31/115 (26%), Positives = 58/115 (49%), Gaps = 6/115 (5%)

Query: 7   TKKKIAKSFKRLFISQAFDKISVSDIMEDAGIRRQTFYNHFVDKYALLEWIFQTELSEQV 66
            TK  +  +   L   Q+F+ ++VSD+ + AGI R TFY H+ DK+ ++   F+ +  + +
Sbjct: 8   TKAYVKTALITLLTEQSFETLTVSDLTKKAGINRGTFYLHYTDKFDMMNH-FKNDTLDDL 66

Query: 67  TDNLD----YISGFQLLSELLTFFKMNQSFYIKLFQIEDQNDFSSYFESYCEQLV 117
               L+    Y    Q+L++ L++    ++EF   L  I     F    + +C Q +
Sbjct: 67  YRLLNQABIYTDTRQVLNQTLSYLIEHRREFITALATI-SYLKFPQLIKDFCYQFL 120
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1815

A DNA sequence (GBSx1922) was identified in *S.agalactiae* <SEQ ID 5643> which encodes the amino acid sequence <SEQ ID 5644>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1974(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1816

A DNA sequence (GBSx1923) was identified in *S.agalactiae* <SEQ ID 5645> which encodes the amino acid sequence <SEQ ID 5646>. This protein is predicted to be dihydroxyacetone kinase (b1200). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1806(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07112 GB:AP001518 dihydroxyacetone kinase [Bacillus halodurans]
 Identities = 141/285 (49%), Positives = 197/285 (68%), Gaps = 1/285 (0%)

Query: 45  IPLLSGGGSGHEPAHFGYVGEGNLSAAISGPIFVPPCASDILETIRFINRGKGVFVIIKN 104
           +PI+SGGGSGHEP H GYVGRGNL+AA+ G +FVPP A  +L  IR +++GKGV +IIKN
Sbjct: 46  VPIISGGGSGHEPGHLGYVGEGNLAAAVHGDVFVPPSAQQVLAAIRQMDQGKGVLLIIKN 105

Query: 105 FEADLEEFSQAIEQARQEGIPIKYIVSHDDISVET-SNFKIRHRGVAGTVLLHKLIGQAA 163
            F ADL  F  A  QAR EG  + +++ +DD+SVE+ ++P+ R RGVAG VL+HKIIG AA
Sbjct: 106 FVADLATFLSAEVQARAEGRDVAHVIVNDDVSVESDASPEKRRRGVAGAVLVHKIIGAAA 165

Query: 164 LEGASLDELEQLGLSLTTSMATLGVASKSATILGQHQPVFDIEEGYISFGIGIHGEPGYR 223
             EG SL+ L++++G  +  ++ATLGVA    A +  + +P F +EEG + FG+GIHGE GYR
Sbjct: 166 KEGYSLSALQEIGEQVVKNLATLGVALTHADLPERREPQFLLEEGEVYFGVGIHGEQGYR 225

Query: 224 TMPFVSMEHLANELVNKLKMKLRWQDGEAPILLINNLGGSSKMEELLFTNAVMEPLALDD 283
              VS E LA ELVNKLK   RW   + + +LIN LGG+  +E+ +F N V   LA+++
Sbjct: 226 KEKLVSSELLAVELVNKLKSLYRWDKNDQYAVLINGLGGTPLIEQYVFANDVRRLLAIEN 285

Query: 284 LQLPPIKTGHLITSLDMAGLSVTLCRVKDSRWIDYLKHKTDARAW 328
            L + F+K G  +TSL+M G+S+T+ ++ D +W+ +L    D  W
Sbjct: 286 LHVSFVKVGTQLTSLNMKGISLTMLKICDEQWVKWLYAPVDVAEW 330
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1817

A DNA sequence (GBSx1924) was identified in *S.agalactiae* <SEQ ID 5647> which encodes the amino acid sequence <SEQ ID 5648>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3902(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10085> which encodes amino acid sequence <SEQ ID 10086> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC75047 GB:AE000290 orf, hypothetical protein [Escherichia coli K12]
 Identities = 182/237 (76%), Positives = 201/237 (84%)

Query: 20  MGRKWANIVAKKTAKDGANSKVYAKPGVEIYVAAKQGEPDPESNSALKFVLDRAKQAQVP 79
           MGRKWANIVAKKTAKDGA SK+YAKPGVEIY AAKQGEPDPE N++LKFV++RAKQAQVP
Sbjct: 1   MGRKWANIVAKKTAKDGATSKIYAKPGVEIYAAAKQGEPDPELNTSLKFVIERAKQAQVP 60

Query: 80  KHVIDKAIDKAKGNTDETFVEGRYEGPGNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN 139
           KHVIDKAIDKAKG  DETFV+GRYEGPGNGSMII +TLTSNVNRT ANVRT + K GGN
Sbjct: 61  KHVIDKAIDKAKGGSDETFVQGRYEGPGNGSMIIAETLTSNVNRTIANVRTAPNKKGGN 120

Query: 140 MGASGSVSYLFDKKGVIVFAGDDADTVPEQLLEADVDVDDVEAERGTTIVYTAPTDLEKG 199
           +GA+GSVSY+FD  GVIVF G D D +PE LLEA+VDV DV  ERG I +YT PTDLEKG
Sbjct: 121 IGAAGSVSYMFDNTGVIVFKGTDPDHIPELLEAEVDVRDVTEERGNIVIYTEPTDLEKG 180

Query: 200 IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVAD 256
           I AL+  G+ EF  TELEMI QSEV L  +DLE FE L+DALE DDDVQKVYHNVA+
Sbjct: 181 IAALKAAGITEFSTTELEMIAQSEVKLSPEDLEIFEGLVDALEDDDDVQKVYHNVAN 237
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5649> which encodes the amino acid sequence <SEQ ID 5650>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2926(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 233/238 (97%), Positives = 236/238 (98%)

Query: 20  MGRKWANIVAKKTAKDGANSKVYAKPGVEIYVAAKQGEPDPESNSALKFVLDRAKQAQVP 79
           MGRKWANIVAKKTAKDGA SKVYAKPGVEIYVAAKQGEPDPE N+ALKFV+DRAKQAQVP
Sbjct: 1   MGRKWANIVAKKTAKDGATSKVYAKPGVEIYVAAKQGEPDPELNTALKFVIDRAKQAQVP 60

Query: 80  KHVIDKAIDKAKGNTDETFVEGRYEGPGNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN 139
           KHVIDKAIDKAKGNTDETFVEGRYEGPGNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN
Sbjct: 61  KHVIDKAIDKAKGNTDETFVEGRYEGPGNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN 120

Query: 140 MGASGSVSYLFDKKGVIVFAGDDADTVPEQLLEADVDVDDVEAERGTTIVYTAPTDLEKG 199
           MGASGSVSYLFDKKGVIVFAGDDAD+VPEQLLEADVDVDDVEAERGTTIVYTAPTDLEKG
Sbjct: 121 MGASGSVSYLFDKKGVIVFAGDDADSVPEQLLEADVDVDDVEAERGTTIVYTAPTDLEKG 180

Query: 200 IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVADF 257
           IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVADF
Sbjct: 181 IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVADF 238
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1818

A DNA sequence (GBSx1925) was identified in *S.agalactiae* <SEQ ID 5651> which encodes the amino acid sequence <SEQ ID 5652>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2507(Affirmative) < succ>
```

```
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1819

A DNA sequence (GBSx1926) was identified in *S.agalactiae* <SEQ ID 5653> which encodes the amino acid sequence <SEQ ID 5654>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1523(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA20826 GB:AL031541 hypothetical protein SCI35.37 [Streptomyces
        coelicolor A3(2)]
 Identities = 73/178 (41%), Positives = 101/178 (56%), Gaps = 2/178 (1%)

Query: 35  VKNAGGLPVILPISEAESAKAYVEMIDKLIISGGQNVLPSYYGEKKIESDDYSLARDIF 94
           V+ AGGL +LP    E A A V  +D ++I+GG +V P  YG E    +  + ARD +
Sbjct: 37  VQRAGGLAAMLPPDAPEHAAATVARVDGVVIAGGPDVEPVRYGAEPDPRTGPPARARDIW 96

Query: 95  EFALVEEALKQNKPIFAICRGMQLVNVALGGTLNQSIDNHYQEPYIGFAHYLNVEKGSFL 154
           E AL+E AL     P+  ICRGMQL+NVALGGTL Q I+ H +    +  H +    G+
Sbjct: 97  ELALIEAALAARVPLLGICRGMQLLNVALGGTLVQHIERHAEVVGVFGHPVRPVPGTLY 156

Query: 155 EGFISGDFKINSLHRQSVKLLAESLIVSARDPRDGTVEAYESRT-EQCIIGVQWHPEL 211
            G +   +  + H Q+V  L  GL+ SA    DGTVEA E  +    ++GVQWHPE+
Sbjct: 157 AGAVPEETFVPTYHHQAVDRLSSGLVASAH-AADGTVEALEMPSSSGWVLGVQWHPEM 213
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5655> which encodes the amino acid sequence <SEQ ID 5656>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1210(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 111/230 (48%), Positives = 145/230 (62%), Gaps = 3/230 (1%)

Query: 2   LTKPIIGITGNERENSDIPGYYYDSVSKHISRGVKNAGGLPVILPISEARSAKAYVEMID 61
           +TKPIIGIT N+R     +     + V  +GGLP++LPI +  +AK YV M+D
Sbjct: 1   MTKPIIGITANQRLNMALDNLPWSYAPTGFVQAVTQSGGLPLLLPIGDEAAAKTYVSMVD 60

Query: 62  KLIISGGQNVLPSYYGEEKIESDDYSLARDIFEFALVEEALKQNKPIFAICRGMQLVNV 121
           K+I+ GGQNV P YY EEK    DD+S  RD FE A+++EA+    KPI  ICRG QL+NV
Sbjct: 61  KIILIGGQNVDPKYYQEEKAAPDDDFSPERDTFELAIIKEAITLKKPILGICRGTQLMNV 120
```

```
Query: 122 ALGGTLNQSIDNHYQE-PYIGFAHYLNVEKGSFLEGFISGDFKINSLERQSVKLLAEGLI 180
           ALGG LNQ ID+H+QE P     +H + +E  S L           INS HRQS+K +A+ L
Sbjct: 121 ALGGNLNQHIDSHWQEAPSDPLSHSMIIEPDSILYPIYGHKTLINSPHRQSLKTVAKDLK 180

Query: 181 VSARDPRDGTVEAYRSRTEQC-IIGVQWHPELMLH-QIENQTLPGYFVNR 228
           V ARDPRDGT+EA  S  +     +GVQWHPEL+   + E+  LF  FVN+
Sbjct: 181 VIARDPRDGTIEAVISTNDAIPFLGVQWHPELLQGVRDEDLQLPRLFVND 230
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1820

A DNA sequence (GBSx1927) was identified in *S.agalactiae* <SEQ ID 5657> which encodes the amino acid sequence <SEQ ID 5658>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5794(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1821

A DNA sequence (GBSx1928) was identified in *S.agalactiae* <SEQ ID 5659> which encodes the amino acid sequence <SEQ ID 5660>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0524(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8905> which encodes amino acid sequence <SEQ ID 8906> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 22   Cremd: 4
McG: Discrim Score:      8.37
GvH: Signal Score (-7.5): -0.64
     Possible site: 21
>>> May be a lipoprotein
ALOM program   count: 0 value:   6.74 threshold:  0.0
   PERIPHERAL  Likelihood =  6.74      112
 modified ALOM score:  -1.85

*** Reasoning Step: 3

----- Final Results -----
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear)
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2919> which encodes the amino acid sequence <SEQ ID 2920>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> May be a lipoprotein

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 120/162 (74%), Positives = 141/162 (86%), Gaps = 5/162 (3%)

Query: 6   LAACSSKSHTTKTGK----KEVNFATVGTTAPFSYVKDGKLTGFDIEVAKAVFKGSDNYK 61
            LAAC S S T ++G     KEV FATVGTTAPFSY K G+LTG+DIEVAKAVFKGSD+YK
Sbjct: 20  LAACGS-SKTAESGNQGSSKEVLFATVGTTAPFSYEKGGQLAGYDIEVAKAVFKGSDDYK 78

Query: 62  VTFKKTEWSSVFTGIDSGKFQMGSNNISYSSSRSQKVLFSYPIGSTPSVLAVPKNSNIKA 121
            V+FKKTEWSS+FTG+DSGK+QMGSNNIS++ ERS KVLFSYPIGSTPSVL VPK+S+IK+
Sbjct: 79  VSFKKTEWSSIFTGLDSGKYQMGSNNISPTKERSAKYLFSYPIGSTPSVLVVPKDSDIKS 138

Query: 122 YNDISGHKTQVVQGTTTAKQLENPNKEHQKNPVTLKYTNENL 163
            ++DI GH TQVVQGTT+  QLE+PNK+H  NPVTLK+TNEN+
Sbjct: 139 FDDIQGHTTQVVQGTTSVAQLEDPNKKHSDNPVTLKFTNENI 180
```

SEQ ID 8906 (GBS71) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 4; MW 31.8kDa).

GBS71-His was purified as shown in Figure 196, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1822

A DNA sequence (GBSx1929) was identified in *S.agalactiae* <SEQ ID 5661> which encodes the amino acid sequence <SEQ ID 5662>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2179(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 2920:

```
Identities = 64/91 (70%), Positives = 78/91 (85%)

Query: 1   MSDGKADFKLPDGPTVNAIIKNQGLTNLKTIPLTMRDQPYIYFIPGQDQKDLQKYVNNRL 60
           +S+GKADFK+PD PTVNAIIKNQGL NLKTI LT  +QP+IYFIP QDQ+ LQ +VN R+
Sbjct: 187 LSEGKADFKIPDAPTVNAIIKNQGLDNLKTIELTSTEQPFIYFIPSQDQEKLQSFVNKRI 246

Query: 61  KQLRKDGTLSKIAKEYLGKDYVPNEKDLVTP 91
           K+L  DGTLSK+AKE+LGKDYVP++K+L  P
Sbjct: 247 KELTADGTLSKLAKEHLGKDYVPSDKELKLP 277
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1823

A DNA sequence (GBSx1930) was identified in *S.agalactiae* <SEQ ID 5663> which encodes the amino acid sequence <*SEQ ID 5664*>. This protein is predicted to be 28 kDa outer membrane protein (yaeC). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.44   Transmembrane   25 -  41 (  25 -  42)

----- Final Results -----
                bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB59825 GB:AJ012388 hypothetical protein [Lactococcus lactis]
 Identities = 110/283 (38%), Positives = 175/283 (60%), Gaps = 13/283 (4%)

Query: 22  KLKHIVLGLALTTLLGV----TPSNQEVSASSTSSKVVKVGVMTFSDTEKARWDKIEKLV 77
           K ++I++ +A+  L+ +     + ++Q    +S   K VKVG+M+     ++  W  +
Sbjct: 4   KNRNIIIAVAVLILVALVAPFSLNHQGGVKAAGAEKTVKVGIMSGDKQDQSVWKSVANTA 63

Query: 78  GDK--AKIKFTEFTDVTQPNQATANKDVDINARQHVNFLMWNKENKKNLIPLEKTYLAP 135
           +K   K+KF  F+DY QPN+A  + D+DINAFQ YN++ WNK +K +++ +  TY+ P
Sbjct: 64  KEKYDLKLKFVYFSDYNQPNEALLSGDIDINAFQSYNYVKTWNKAHKSDIVAVGNTYITP 123

Query: 136 IRIYSEKVKSLKKLKKGATIAIPNDATNGSRALYVLQSAGLIKLNVS-GKKVATVANITS 194
           + IYS+++  L  LK+G+T+AIPNDA+N SRALFVLQSAGL+KL  S   K+  + +IT
Sbjct: 124 MHIYSKEISKLSDLKEGSTVAIPNDASNESRALFVLQSAGLLKLTTSDSSKLVGLPDITE 183

Query: 195 NKKDINIQELDASQTPRALKDVDAAIINNTYIEQANLKPSDAIPVEKSDKNSKQWINIIA 254
           N   + +E+DASQTPRAL  V  +++N  Y   A+L  S+++F+E  +K S Q+IN IA
Sbjct: 184 NPHQLKFKEVDASQTPRALDSVALSVVNYNYATAASLPKSESVFMEPLNKTSAQYINFIA 243

Query: 255 GRKNWKKQKNAKAIQAILDAYHTDEVKKVIKDTSAD---IPQW 294
           G+KN K  + +  AY +   +K IK+   D   +P W
Sbjct: 244 ---TTSKEKNNKVYKEVAKAYASKATEKAIKEQYPDGGELPAW 283
```

There is also homology to SEQ ID 2132.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8907> and protein <SEQ ID 8908> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:       7.47
GvH: Signal Score (-7.5): -4.79
     Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -1.44 threshold:  0.0
   INTEGRAL    Likelihood = -1.44   Transmembrane    5 -  21 (   5 -  22)
   PERIPHERAL  Likelihood =  5.20       147
 modified ALOM score:   0.79

*** Reasoning Step: 3

----- Final Results -----
```

```
bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
40.6/63.1% over 279aa
                                                          Lactococcus lactis
  GP|6165402| hypothetical protein Insert characterized

ORF00442(364 - 1182 of 1482)
GP|6165402|emb|CAB59825.1|AJ012388(4 - 283 of 287) hypothetical protein (Lactococcus
lactis)
%Match = 21.0
%Identity = 40.6  %Similarity = 63.0
Matches = 112  Mismatches = 96  Conservative Sub.s = 62

162       192       222       252       282       312       342       372
WDTFKNS*RIFWR*LRTK*ERSRYS*GEVVIKTKEMSILSFLLYSLKL*QETVYNNLILITSYGIISLSQKLREPIMKLK
                                                                          | :
                                                                          MNFKNR

402       450       480       510       540       564       594
HIVLGLALTTLLG--VTFS--NQKVSASSTSSKVVKVGVMTFSDTEKARWDKIRKLVGDK--AKIKFTEFTDYTQPWQAT
 :|:: :|:  |:   || :|    :|   | ||||:|:  ::  |  :    :|  :||  |:|| |||:|
NIIIAVAVLLILVALVAFFSLNHQGGVKASAGEKTVKVGIMSGDKQDQEVWKSVANTAKEKYDLKLKFVYFSDYNQPNEAL
                  20        30        40        50        60        70        80

624       654       684       714       744       774       804       834
ANKDVDINAFQHYNFLENWNKENKKNLIPLEKTYLAPIRIYSSKVKSLKKLKKGATIAIPNDATNGSSRALYVLQSAGLIK
 : |:|||||| ||::: ||| :| ::: :  ||: |: |||:::  |  ||:|:|:|||||:| ||||:||||||:|
LSGDIDINAFQSYNYVKTWNKAHKSDIVAVGNTYITPMHIYSKEISKLSDLKEGSTVAIPNDASNSSRALFVLQSAGLLK
                  100       110       120       130       140       150       160

861       891       921       951       981       1011      1041      1071
LNVS-GKKVATVANITSNKKDINIQELDASQTPRALKDVDAAIINNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRK
|   |    |:  : :|| |   :  :|:|||||||||  |  :::|  |   |:|  |:::|:|  :|  | |:|| ||
LTTSDSSKLVGLPDITENPHQLKFKEVDASQTPRALDSVALSVVNYNYATAASLPKSESVFMEPLNKTSAQYINFIA---
                  180       190       200       210       220       230       240

1101      1131      1161      1182      1212      1242      1272      1302
NWKKQKNAKAIQAILDAYHTDEVKKVIKDTSAD---IPQW*RELTV*V*QGILIGYNLSAI*P*RAWDEYNVPGSWIVFE
  .  |:||  |  : :  || :   :|  ||:   |   :| |  :|
TTSKEKNNKVYKEVAKAYASKATEKAIKEQYPDGGELPAWDLKL
                  260       270       280
```

SEQ ID 8908 (GBS35) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 2; MW 31.6kDa).

The GBS35-His fusion product was purified (Figure 96A; see also Figure 192, lane 6) and used to immunise mice (lane 2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 96B), FACS (Figure 96C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Example 1824

A DNA sequence (GBSx1931) was identified in *S.agalactiae* <SEQ ID 5665> which encodes the amino acid sequence <SEQ ID 5666>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3126(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF11560 GB:AE002038 ArgE/DapE/Acyl family protein [Deinococcus radiodurans]
 Identities = 129/419 (30%), Positives = 210/419 (49%), Gaps = 14/419 (3%)

Query: 26  LRDLIAIKSIFAQKVGLNDLSSYLGEVPIKAGAEVIIDDSYSGAPFIVANFKSSKVDAKRI 85
           LR L+A+ S+ AQ   L + +  +  +    G V       AP ++A            +
Sbjct: 16  LRALVALPSVSAQGRMLPETADAVAGLLRABGFGVQQFPGTVAPVLLAEAGEGPPT----L 72

Query: 86  IFYNHYDTVPADEVEQWTEDPFTLSLRYGKMYGRGVDDDKGHITARLSAVKKYLKRHKGE 145
           + YNHYD  P D +E W   PF L+ R G++YGRG  DDKG + +RL+AV+  +    G
Sbjct: 73  LIYNHYDVQPEDPLELWDTPPFELTERGGHLYGRGASDDKGELASRLAAVRA-VREQLGH 131

Query: 146 LPLDITFIVEGAEESGSVGLDYYLEKYQEQLQADLIVWEDGPKNFKGQLEIAGCNKGIV 205
           LP+ I +++EG EE  S  L+ ++ ++  +LQ AD   WE G  +P+G+  ++ G KG++
Sbjct: 132 LPVKIKWLIEGEEEVGSPTLERFVAEHAAELQ-ADGCWWEPGGISPEGRPILSLGLKGVM 190

Query: 206 TFDLSVSSADVDIHSSFGGVVDSSTWYLIQALNTLRDNKGHILVEGIYDKVIPPTKRELE 265
             +L    AD D+HSS G V+D+  + L +A+ +LRD +G++ + G YD V   +  + +
Sbjct: 191 CLELRCRVADSDLHSSLGAVIDNPLYCLARAVASLRDEQGNVTIPGFYDDVRAASGADRQ 250

Query: 266 LVEKYSYRSAVALEGAYQLVLPSLADSHKTFLRKLYFEPSIAIEGITSGYQGEGVKTILP 325
           + +   +A+  + +  P   +  +  P + + G  GYQSEG KT+LP
Sbjct: 251 AIAQIP-GDGQAVRDTFGVRRP--LATGPAYNERTNLHPVVNVNGWGGGYQGESSKTVLP 307

Query: 326 AYAKCKAEVRLVPGLTPKGVLDSIQNHLKKNGFKDIELT-YTLSEMSYRSDMSAPSILKV 384
             K + RLVP  P VL ++ HL  G DIE+    +  R+D  P +
Sbjct: 308 GAGPVKLDFRLVPDQDPARVLSLLREHLTAQGLSDIEVVELEAHQKPARADAGHPFVQAC 367

Query: 385 VDLAEQFYPEGISLLPTSPGTGPMY-----LVHQALRAPIAAIGIGHANSRDHGVDENV 438
           V A   + +  P+S  +GPM+          L . P  A+GIG+   R H  +EN+
Sbjct: 368 VAAARAAHGQDPIVHPSSGASGPMFPFTGGAGGGGLGIPCVAVGIGNHAGRVHAPNENI 426
```

There is also homology to SEQ ID 2588.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1825

A DNA sequence (GBSx1932) was identified in *S.agalactiae* <SEQ ID 5667> which encodes the amino acid sequence <SEQ ID 5668>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5366(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB59828 GB:AJ012388 hypothetical protein [Lactococcus lactis]
 Identities = 187/338 (55%), Positives = 256/338 (75%), Gaps = 12/338 (3%)

Query: 6   IIKLDNIDVTFHQKKREINAVKDVTIHINQGDIYGIVGYSGAGKSTLVRVINLLQEPSAG 65
           II+L+N+ V FHQK R + AVK+ T+HI +GDIYG++GYSGAGKSTLVR INLLQ+P+ G
Sbjct: 4   IIELNNLSVQFHQGRLVTAVKNATLHIEKGDIYGVIGYSGAGKSTLVRTINLLQKPTEG 63

Query: 66  KITIDDQVIYD--NKVTLTSTQLREQRREIGMIFQHFNLMSQLTAEQNVAFALKHSG--- 120
           +I I+ + I+D  N V  T  +LRE R++IGMIFQHFNL+S+ T   NVAFALKHS
```

```
Sbjct: 64  QIVINGEKIFDSKNPVKFTGAKLREFRQKIGMIFQHFNLLSEKTVFNNVAFALQHSQIED 123

Query: 121 -------LSKEAKAAKVAKLLELVGLSDRAQNYPSQLSGGQKQRVAIARALANDPKILIS 173
                 L+K+ K  KV +LL+LV L+D +  YP+QLSGGQKQRVAIARALANDP+ILIS
Sbjct: 124 KNGKKRYLVTKKEKNDKVTELLKLVDLADLSDKYPAQLSGGQKQRVAIARALANDPEILIS 183

Query: 174 DESTSALDPKTTKQILALLQDLNKKLGLTIVLITHEMQIVKDIANRVAVMQNGKLIEEGS 233
           DE TSALDPKTT QIL LL+ KLG+T+VLITHEMQ+VK+IAN+VAVMQNG++IE+ S
Sbjct: 184 DEGTSALDPKTTNQILDLLKSLHEKLGITVVLITHEMQVVKEIANKVAVMQNGEIIEQNS 243

Query: 234 VLDIFSHPRESLTQDFIKIATGIDEAMLKIEQQEVVKNLPVGSKLVQLKVAGHSTDEPLL 293
           ++DIF+ P+E+LT+ FI+  + ++  +  + E++  L    +L+ L Y+G  ++P++
Sbjct: 244 LIDIFAQPKEALTKQFIETTSSVNRFIASLSKTSLLAQLADDEELIHLDYSGSEELRDPVV 303

Query: 294 NQIYKEFEVTANILVGNIEILDGIPVGEMVVILSGDEE 331
           + I K+F+VT NI YGN+E+L G P G +V+ L G  E
Sbjct: 304 SDITKKFDVTTNIFYGNVELLQGQPFGSLVLTLKGSSE 341
```

There is also homology to SEQ ID 76.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1826

A DNA sequence (GBSx1933) was identified in *S.agalactiae* <SEQ ID 5669> which encodes the amino acid sequence <SEQ ID 5670>. This protein is predicted to be ABC transporter, permease protein. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood =-12.79   Transmembrane  203 - 219 ( 197 - 225)
   INTEGRAL     Likelihood = -8.86   Transmembrane   73 -  89 (  69 - 102)
   INTEGRAL     Likelihood = -7.38   Transmembrane   38 -  54 (  35 -  56)
   INTEGRAL     Likelihood = -1.12   Transmembrane  103 - 119 ( 103 - 119)

----- Final Results -----
                bacterial membrane --- Certainty=0.6116(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10083> which encodes amino acid sequence <SEQ ID 10084> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB59829 GB:AJ012388 hypothetical protein [Lactococcus lactis]
 Identities = 137/231 (59%), Positives = 171/231 (73%), Gaps = 1/231 (0%)

Query: 1   MIEWIQTHLPNVYQMGWEGAYGWCTAIVQTLYMTFWSFLIGGLMGLLGGLFLVLTSPRGV 60
           M EW     PNV  +GW G  GW TAIVQTLYMTF S LIGGL+GL+ G+ +V+T+  G+
Sbjct: 1   MAEWFAHTFPNVVYLGWTGETGWWTAIVQTLYMTFISALIGGLLGLIPGIGVVVTAEDGI 60

Query: 61  IANKLVPGVLDKVVSVFRALPPIILLALIAPVTRVIVGTTLGSPAALVPLSLAVFPFFAR 120
           N+ +F +LDK+VS+ RA PFIILLA IAP+T+++VGT +G  AALVPL+L V PF+AR
Sbjct: 61  TPNRPLFWILDKIVSIGRAFPFIILLAAIAPLTKILVGTQIGVTAALVPLAIGVAPFYAR 120

Query: 121 QVQVVLAELDGGVIEAAQASGGTLWDII-VVYLRELGLPDLIRVSTVTLISLVGETAMAGA 179
           QVQ  L  +D G +EAAQ  G    DI+  VYLRE L  LIRVSTVTLISL+G TAMAGA
Sbjct: 121 QVQASLESVDHGKVEAAQ-VGADFLDIVF-VYLRELASLIRVSTVTLISLIGLTAMAGA 180

Query: 180 IGAGGLGSVAITKGYNYSRDDTTLVATILILLLIFFIQFLGDFLTRRLSHK 230
           IGAGGLG+ AI+ GYN    +D+T  ATILLL+ +   +Q +GDFL RR+SH+
Sbjct: 181 IGAGGLGNTAISYGYNRFANDVTWFATILILIFVLLVQLVGDFLARRVSHR 231
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5671> which encodes the amino acid sequence <SEQ ID 5672>. Analysis of this protein sequence reveals the following:

```
        Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.15   Transmembrane  194 - 210 ( 187 - 215)
   INTEGRAL    Likelihood =-10.67   Transmembrane   28 -  44 (  20 -  52)
   INTEGRAL    Likelihood = -8.12   Transmembrane   70 -  86 (  62 -  91)

----- Final Results -----
                bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB59829 GB:AJ012388 hypothetical protein [Lactococcus lactis]
 Identities = 123/213 (57%), Positives = 153/213 (71%), Gaps = 1/213 (0%)

Query: 9   GDAGWGLAIWNTLYMTIVPFIVGSAIGLLLGLLLVLTGPDGVIENKTICWVIDKVTSIFR 68
           G+ GW  AI  TLYMT +  ++GG +GL+ G+ +V+T  DG+  N+ + W++DK+ SI R
Sbjct: 19  GETGWWTAIVQTLYMTFISALIGGLLGLIFGIGVVVTAEDGITPNRPLFWILDKIVSIGR 78

Query: 69  AIPFVILIAILASFTVLLRTTLGATAALVPLTFATFPFYARQVQVVFSELDKGVIEAAQ 128
           A PF+IL+A +A  T +L+ T +G TAALVPL      PFYARQVQ     +D G +EAAQ
Sbjct: 79  APPFIILLAAIAPLTKILVGTQIGVTAALVPLALGVAPFYARQVQASLESVDHGKVEAAQ 138

Query: 129 ASGATFWDIV-KVYLSEGLPDLIRVSTVTLISLVGETAMAGAIGAGGLGNVAISYGYNRF 187
             GA F DIV  VYL E L  LIRVSTVTLISL+G TAMAGAIGAGGLGN AISYGYNRF
Sbjct: 139 TVGADFLDIVFTVYLREELASLIRVSTVTLISLIGLTAMAGAIGAGGLGNTAISYGYNRF 198

Query: 188 NNDVTWVATIILLIIFAIQFIGDSLTRRFSHK 220
           NDVTW ATI+LL+ +Q +GD L RR SH+
Sbjct: 199 ANDVTWFATILLIFVLLVQLVGDFLARRVSHR 231
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 146/212 (68%), Positives = 172/212 (80%)

Query: 19  GAYGWQTAIVQTLYMTFWSFLIGGLMGLLGGLFLVLTSPRGVIANKLVFGVLDKVVSVFR 78
           G GW  AI  TLYMT   F++GG +GLL GL LVLT P GVI NK +  V+DKV S+FR
Sbjct: 9   GDAGWGLAIWNTLYMTIVPFIVGSAIGLLLGLLLVLTGPDGVIENKTICWVIDKVTSIFR 68

Query: 79  ALPFIILLALIAPVTRVIVGTTLGSPAALVPLSLAVFPFFARQVQVVLAELDGGVIRAAQ 138
           A+PF+IL+A++A  T +++ TTLG+ AALVPL+ A FPF+ARQVQVV +ELD GVIEAAQ
Sbjct: 69  AIPFVILIAILASFTVLLRTTLGATAALVPLTFATFPFYARQVQVVFSELDKGVIEAAQ 128

Query: 139 ASGGTLWDIVVVLRGELPDLIRVSTVTLISLVGETAMAGAIGAGGLGSVAITKYGYNYSR 198
           ASG T WDI+ VYL RGLPDLIRVSTVTLISLVGETAMAGAIGAGGLG+VAI+ GYN
Sbjct: 129 ASGATFWDIVKVYLSEGLPDLIRVSTVTLISLVGETAMAGAIGAGGLGNVAISYGYNRFN 188

Query: 199 DDITLVATILILLLIFFIQFLGDFLTRRLSHK 230
           +D+T VATI+LLL IF IQF+GD LTRR SHK
Sbjct: 189 NDVTWVATIILLLIFAIQFIGDSLTRRFSHK 220
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1827

A DNA sequence (GBSx1934) was identified in *S.agalactiae* <SEQ ID 5673> which encodes the amino acid sequence <SEQ ID 5674>. This protein is predicted to be alcohol dehydrogenase, zinc-containing (Zn-dependent). Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -2.92   Transmembrane   71 -  87 (  69 -  87)

----- Final Results -----
                bacterial membrane --- Certainty=0.2168(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9419> which encodes amino acid sequence <SEQ ID 9420> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF41759 GB:AE002488 alcohol dehydrogenase, zinc-containing
          [Neisseria meningitidis MC58]
 Identities = 135/246 (54%), Positives = 186/246 (74%), Gaps = 1/246 (0%)

Query: 3   SHCEDGGWILGHLIEGTQAEYVHIPHADGSLYHAPEGVCDDALVMLSDILPTSYEIGVLP 62
           SHC +GGWILG++I+GTQAEYV  P+AD SL    P+ V ++  ++LSD LPT++EIGV
Sbjct: 102 SHCRNGGWILGYMIDGTQAEYVRTPYADNSLVPLPDNVNESIALLLSDALPTAHEIGVQY 161

Query: 63  SHIKPGDTVCIVGAGPIGLSALLTAQFYSPAKIIMVDLSQKRLEASKKFGATNTILSTST 122
            +KPGDTV I GAGP+G+SALLTAQ YSPA II+ D+ + RL+  +K+ GATHTI + ++
Sbjct: 162 GDVKPGDTVFIAGAGPVGMSALLTAQLYSPAAIIVCDMDENRLKLAKELGATHTI-NPAS 220

Query: 123 QEVKEEIDKITKGRGVDVVLECVGYPATPDICQNVVSIGGHIANVGVHGKPVEFNLQDLW 182
            EV +++  I    GVD  +E VG PAT+++CQ++V  GGHIA VGVHG+ V+F L+ LW
Sbjct: 221 GEVSKQVFAIVGEDGVDCAIEAVGIPATWNMCQDIVKPGGHIAVVGVHGQSVDFKLEKLW 280

Query: 183 IKNITLNTGLVNANTTEMLLEVLETGKIDATQLVTHHFKLSEIEEAYKVFKAAEENNTLK 242
           IK + + TGLVNANTTEML++ + +  +D T+++THHFK SE+E+AY VFK A EN  +K
Sbjct: 281 IKKLAITTGLVNANTTEMLMKAISSSSVDYTKMLTHHFKFSELEKAYDVFKHAAENQVMK 340

Query: 243 VIIEND 248
           V++E D
Sbjct: 341 VVLEAD 346
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 785> which encodes the amino acid sequence <SEQ ID 786>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -5.41   Transmembrane   184 - 200 ( 183 - 203)

----- Final Results -----
                bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 199/250 (79%), Positives = 226/250 (89%)

Query: 1   MPSHCEDGGWILGHLIEGTQAEYVHIPHADGSLYHAPEGVCDDALVMLSDILPTSYEIGV 60
           + SHC+DGGWILGHLI GTQAEYVHIPHADGSLYHAP+ + D+ALVMLSDILPTSYEIGV
Sbjct: 114 LSSHCQDGGWILGHLINGTQAEYVHIPHADGSLYHAPDTIDDEALVMLSDILPTSYEIGV 173

Query: 61  LPSHIKPGDTVCIVGAGPIGLSALLTAQFYSPAKIIMVDLSQKRLEASKKFGATNTILST 120
           LPSH+KPGD VCIVGAGP+GL+ALLT QF+SPA IIMVDLSQ RLEA+K FGATHTI S
Sbjct: 174 LPSHVKPGDNVCIVGAGPVGLAALLTVQFFSPANIIMVDLSQNRLEAAKTFGATHTICSG 233

Query: 121 STQEVKEEIDKITKGRGVDVVLECVGYPATPDICQNVVSIGGHIANVGVHGKPVEFNLQD 180
           S++EVK  ID IT GRGVD+ +ECVGYPATPDICQ ++S+GGHIANVGVHGKPV+FNL +
Sbjct: 234 SSKEVKAIIDDITNGRGVDISMECVGYPATPDICQKIISVGGHIANVGVHGKPVDFNLDE 293
```

```
Query: 181 LNIKNITLNTGLVNANTTEMLLEVLETGKIDATQLVTHHPKLSEIEEAYKVFKAAEENNT 240
           LNIKNITLNTGLVNANTTEMLL VL+TGKIDAT+L+THHPKLSE+E+AY+ FK A  NN
Sbjct: 294 LNIKNITLNTGLVNANTTEMLLNVLKTGKIDATRLITHHPKLSEVEKAYETFKHAGANNA 353

Query: 241 LKVIIENDIT 250
           LKVII+NDI+
Sbjct: 354 LKVIIDNDIS 363
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1828

A DNA sequence (GBSx1935) was identified in *S.agalactiae* <SEQ ID 5675> which encodes the amino acid sequence <SEQ ID 5676>. This protein is predicted to be a dehydrogenase fragment. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
  INTEGRAL    Likelihood =-10.46   Transmembrane   47 -  63 (  33 -  66)

----- Final Results -----
                bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 786:

```
Identities = 23/38 (60%), Positives = 28/38 (73%)

Query: 7  WRNSNMRAATYLSANELSLMTDKAKPQVIKPTDAVVKLV 44
          ++  NM+AATYLS   L L DK KP +IKPTDA+V LV
Sbjct: 10 YKKLNMKAATYLSTGNLQLIDKPKPVIIKPTDAIVQLV 47
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1829

A DNA sequence (GBSx1936) was identified in *S.agalactiae* <SEQ ID 5677> which encodes the amino acid sequence <SEQ ID 5678>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1001(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1830

A DNA sequence (GBSx1937) was identified in *S.agalactiae* <SEQ ID 5679> which encodes the amino acid sequence <SEQ ID 5680>. This protein is predicted to be branched chain amino acid transport system II carrier protein (brnQ). Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.66   Transmembrane  158 - 174 ( 154 - 177)
   INTEGRAL    Likelihood = -6.64   Transmembrane  233 - 249 ( 231 - 252)
   INTEGRAL    Likelihood = -5.20   Transmembrane   37 -  53 (  30 -  57)
   INTEGRAL    Likelihood = -3.98   Transmembrane   90 - 106 (  87 - 108)
   INTEGRAL    Likelihood = -0.80   Transmembrane  130 - 146 ( 130 - 146)

----- Final Results -----
            bacterial membrane --- Certainty=0.4864(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9417> which encodes amino acid sequence <SEQ ID 9418> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC00400 GB:AF008220 branch-chain amino acid transporter
             [Bacillus subtilis]
 Identities = 89/250 (35%), Positives = 139/250 (55%), Gaps = 18/250 (7%)

Query: 1   MDALASIAFAIIVIQASKQYGAITKKEITSMALKSGAIATFLLAFIYIFVGRIGATSQSL 60
           MDALASI F ++V+ A K G    K + +  +K+G IA   L FIY+ +  +GATS +
Sbjct: 199 MDALASIVFGVVVVNAVKSKGVTQSKALAAACIKAGVIAALGLTFIYVSLAYLGATSTNA 258

Query: 61  FKFANGSFLLHNTPI-DGGHVLSQSANFYLGIVGQAILGTAIPLACLTTATGLITACAEY 119
                    P+ +G  +LS S+++  G +G  +LG AI +ACLTT+ GL+T+C +Y
Sbjct: 259 IG----------PVGEGAKIISASSHYLPGSLGNIVLGAAITVACLTTSIGLVTSCSQY 307

Query: 120 FSKLLPKISHITWATIFTLIAITFYFPGGLSEIIRWSLPVLYLLYPLTIVLIPLVFPDQKF 179
           F KL+P +S+     TI TL ++     GL++II +S+P+P+L  +YPL IV+I L F D+ F
Sbjct: 308 FSKLIPALSYKIVVTIVTLFSLIIANFGLAQIIAFSVPILAIYPLAIVIIVLSFIDKIF 367

Query: 180 ESSRIVVQTSIAATAVAALYDALSKLGEMIGLPTIPSALITFFFTKVVPLGEYSMGWISFA 239
           + R VY   + T + ++ D +   G   G      +L  F    +PL    +GN+
Sbjct: 368 KERREVYIACLIGTGLFSILDGIKAAGPSLG------SLDVFLNANLPLYSLGIGWVLPG 421

Query: 240 ICGVLVGLIL 249
           I G ++G +L
Sbjct: 422 IVGAVIGYVL 431
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2233> which encodes the amino acid sequence <SEQ ID 2234>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.83   Transmembrane  235 - 251 ( 228 - 258)
   INTEGRAL    Likelihood = -8.49   Transmembrane  434 - 450 ( 429 - 454)
   INTEGRAL    Likelihood = -8.12   Transmembrane  359 - 375 ( 356 - 377)
   INTEGRAL    Likelihood = -7.86   Transmembrane  150 - 166 ( 144 - 171)
   INTEGRAL    Likelihood = -6.00   Transmembrane  298 - 314 ( 288 - 316)
   INTEGRAL    Likelihood = -5.95   Transmembrane   42 -  58 (  36 -  63)
   INTEGRAL    Likelihood = -3.35   Transmembrane  336 - 352 ( 335 - 354)
   INTEGRAL    Likelihood = -2.81   Transmembrane  199 - 215 ( 198 - 218)
   INTEGRAL    Likelihood = -2.18   Transmembrane  120 - 136 ( 120 - 138)
   INTEGRAL    Likelihood = -1.81   Transmembrane  390 - 406 ( 390 - 407)
   INTEGRAL    Likelihood = -1.01   Transmembrane   81 -  97 (  81 -  97)
```

```
----- Final Results -----
                bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 161/253 (63%), Positives = 197/253 (77%)

Query: 1   MEALASIAFAIIVIQASKQYGAITKKEITSMALKSGAIATFLLAFIYIFVGRIGATSQSL 60
           MEALAS+ FAI+VI+A+KQ+GA T KE+T + L SGAIA  LLA +YIFVGRIGATSQSL
Sbjct: 202 MEALASLVFAILVIEATKQFGAKTDKEMTKITLISGAIAILLLALVYIFVGRIGATSQSL 261

Query: 61  FKFANGSFLLHNTPIDGHWLSQSANFYLGIVGQALGTAIFLACLTTATGLITACAEYF 120
           F F +GSF LH  P++GG +LS ++ FYLG +GQA L  IFLACLTT+TGLIT+ AEYF
Sbjct: 262 FPFIDGSFTLHGNPVMGGQILSHASRFYLGGIGQAFLAVVIFLACLTTSTGLITSSAEYF 321

Query: 121 HKLLPKISHITWATIFTLIAITFYFGGLSEIIRWSLPVLYLLYPLTIVLIFLVFFDQKPE 180
           HKL+P +SHI WATIFTL++  FYFGGLS II WS PVL+LLYPLT+ LIFLV   + F
Sbjct: 322 HKLVPALSHIAWATIFTLLSAFFYFGGLSVIINWSAPVLFLLYPLTVDLIFLVIAQKCFN 381

Query: 181 SSRIVYQTSIAATAVAALYDALSKLGEMTGLFTIPSALTTFFTKVVPLGEYSMGWISFAI 240
           +  IVY+T+I  T + A++DAL  L +MTGLF +P A+ TFF K VPLG++SMGWI FA
Sbjct: 382 NDPIVYRTTIGLTFIPAIFDALLTLSQMTGLFHLPEAVVTFFQKTVPLGQFSMGWIIFAA 441

Query: 241 CGVLVGLILKKVK 253
            G L+GLIL K K
Sbjct: 442 IGFLIGLILSKTK 454
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1831

A DNA sequence (GBSx1938) was identified in *S.agalactiae* <SEQ ID 5681> which encodes the amino acid sequence <SEQ ID 5682>. This protein is predicted to be 30S ribosomal protein S12 (rpsL). Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3698(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9429> which encodes amino acid sequence <SEQ ID 9430> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA78825 GB:Z15120 ribosomal protein S12 [Streptococcus pneumoniae]
 Identities = 64/71 (90%), Positives = 68/71 (95%)

Query: 1  MPTINQLVRKPRKSKVEKSDSPALNIGYNSHRKVHTKLSAPQKRGVATRVGTMTPKKPNS 60
          MPTINQLVRKPRKSKVEKS SPALN+GYNSH+KV T +S+PQKRGVATRVGTMTPKKPNS
Sbjct: 1  MPTINQLVRKPRKSKVEKSKSPALNVGYNSHKKVQTNVSSPQKRGVATRVGTMTPKKPNS 60

Query: 61 ALRKFARVRLS 71
          ALRKFARVRLS
Sbjct: 61 ALRKFARVRLS 71
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5683> which encodes the amino acid sequence <SEQ ID 5684>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3879(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 44/48 (91%), Positives = 47/48 (97%)

Query: 24 LNIGYNSHRKVHTKLSAPQKRGVATRVGTMTPKKPNSALRKFARVRLS 71
          LNIGYNSH+KV TK++APQKRGVATRVGTMTPKKPNSALRKFARVRLS
Sbjct: 1  LNIGYNSHKKVQTKMAAPQKRGVATRVGTMTPKKPNSALRKFARVRLS 48
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1832

A DNA sequence (GBSx1939) was identified in *S.agalactiae* <SEQ ID 5685> which encodes the amino acid sequence <SEQ ID 5686>. This protein is predicted to be purR. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.37   Transmembrane  142 - 158 ( 142 - 159)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA10902 GB:AJ222642 purR [Lactococcus lactis]
 Identities = 143/269 (53%), Positives = 195/269 (72%), Gaps = 1/269 (0%)

Query: 3   LRRSERMVVISNYLINNPYTLTSLNTPASKYGAAKSSISEDIAIIKKAFEQAQIGDIKTV 62
            ++R+ER+V  +N+LIN+P  + +LN  +  Y  AKSSISED+  IK+ FE   +G ++T
Sbjct: 1   MKRNERLVDPTNFLINHPNQMLNLNELSKHYEVAKSSISEDLVPIKRVPENQGVGLVETF 60

Query: 63  TGASGGVIFTPTIAEAEAKEIVEELRQRLSENDRILPGGYIYLSDLLSTPKMLQSIGRII 122
             G+ GGV FTP I +  + E+ +E+ + L E +RILPGGYIYLSD+L TP  L+ IG+II
Sbjct: 61  PGSLGGVRFTPYITDERSLEMSQEIAELLRHENRILPGGYIYLSDILQTPSNLRKIGQII 120

Query: 123 ANAYRGQKIDAVMTVATKGVPLANAVANVLDVPFVIVRRDLKITEGSTVSVNYASGSSGR 182
            A+ Y  +++D VMT+ATKG+P+A +VA +LDVPFVIVRRD K+TEG+T++VNY SGSS R
Sbjct: 121 AHEYHEKQVDVVMTIATKGIPIAQSVAEILDVPFVIVRRDPKVTEGATLNVNYMSGSSSR 180

Query: 183 IEKMFLSKRSLKPNSRVLIVDDFLKGGGTVSGMISLLSEFPDSTLWGVAVFAENA-QEQRE 241
            +E M LSKRSL      VLIVDDF+KG GT++GM SL+ EFD  L GVAVF E    + +R
Sbjct: 181 VENMTLSKRSLSIGQNVLIVDDFMKGAGTINGMRSLVHEFDCLLAGVAVFLKGPFKGERL 240

Query: 242 KMAYKSLLRVSEIDVKNNRVSVEAGNIFD 270
              YKS+L+V  ID+ N  + V+ GNIF+
Sbjct: 241 IDDYKSILKVDRIDIANRSIDVQLGNIFN 269
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5687> which encodes the amino acid sequence <SEQ ID 5688>. Analysis of this protein sequence reveals the following:

```
      Possible site: 41
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL     Likelihood = -1.97   Transmembrane  142 - 158 ( 142 - 160)

   ----- Final Results -----
                  bacterial membrane --- Certainty=0.1786(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA10902 GB:AJ222642 purR [Lactococcus lactis]
 Identities = 142/269 (52%), Positives = 196/269 (72%), Gaps = 1/269 (0%)

Query: 3   LRRSERMVVISNYLINNPYKLTSLNTFATKYEAAKSSISEDIAIIKKAFEEANIGDIDTL 62
           ++R+ER+V  +N+LIN+P ++ +LN  +  YE AKSSISED+  IK+ FE   +G ++T
Sbjct: 1   MKRNERLVDFTNFLINHPNQMLNLNELSKHYEVAKSSISEDLVFIKRVFENQGVGLVETF 60

Query: 63  TGASGGVIFTPSISETEARTIVEDLCQRLSESDRILPGGYIYLSDLLSTPKILQNIGRII 122
            G+ GGV FTP I++  +  + +++ + L E +RILPGGYIYLSD+L TP  L+ IG+II
Sbjct: 61  PGSLGGVRFTPYITDERSLEMSQEIAELLREENRILPGGYIYLSDILGTPSNLRKIGQII 120

Query: 123 ANAFKGEKIDAVMTVATKGVPLANAVANILSVPFVIVRRDLKITEGSTVSVNYASASSDR 182
           A+ +  +++D VMT+ATKG+P+A +VA IL VPFVIVRRD K+TEG+T++++VNY S SS R
Sbjct: 121 AHEYHEKQVDVVMTIATKGIPIAQSVAEILDVPFVIVRRDPKVTEGATLNVNYMSGSSSR 180

Query: 183 IEKMFLSKRSLKPNSRVLIVDDPLKGGGTITGMISLLTEPDSTLVGVAVFAENA-QSERE 241
           +E M LSKRSL     VLIVDDP+KG GTI GM SL+ EPD  L GVAVF E   + ER
Sbjct: 181 VENMTLSKRSLSIQQNVLIVDDPMKGAGTINGMRSLVHEPDCLLAGVAVFLEGPFKGERL 240

Query: 242 QMTFKSLLKVSEIDVKNNVVVEVGNIFD 270
           +KS+LKV  ID+ N ++ V++GNIF+
Sbjct: 241 IDDYKSILKVDRIDIANRSIDVQLGNIFN 269
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 234/270 (86%), Positives = 255/270 (93%)

Query: 1   MKLRRSERMVVISNYLINNPYTLTSLNTFASKVGAAKSSISEDIAIIKKAFEQAQIGDIK 60
           MKLRRSERMVVISNYLINNPY LTSLNTFA+KY AAKSSISEDIAIIKKAFE+A IGDI
Sbjct: 1   MKLRRSERMVVISNYLINNPYKLTSLNTFATKYEAAKSSISEDIAIIKKAFEEANIGDID 60

Query: 61  TVTGASGGVIFTPTIAEAEAKEIVEELRQRLSENDRILPGGYIYLSDLLSTPKMLQSIGR 120
           T+TGASGGVIFTP+I+E EA+ IVE+L QRLSE+DRILPGGYIYLSDLLSTPK+LQ+IGR
Sbjct: 61  TLTGASGGVIFTPSISETEARTIVEDLCQRLSESDRILPGGYIYLSDLLSTPKILQNIGR 120

Query: 121 IIANAYRGQKIDAVMIVATKGVPLANAVANVLDVPFVIVRRDLKITEGSTVSVNYASGSS 180
           IIANA++G+KIDAVMTVATKGVPLANAVAN+L VPFVIVRRDLKITEGSTVSVNYAS SS
Sbjct: 121 IIANAFKGEKIDAVMTVATKGVPLANAVANILSVPFVIVRRDLKITEGSTVSVNYASASS 180

Query: 181 GRIEKMFLSKRSLKPNSRVLIVDDPLKGGGTVSGMISLLSEPDSTLVGVAVFAENAQSQR 240
            RIEKMFLSKRSLKPNSRVLIVDDPLKGGGT++GMISLL+EPDSTLVGVAVFAENAQ +R
Sbjct: 181 DRIEKMFLSKRSLKPNSRVLIVDDPLKGGGTITTGMISLLTEPDSTLVGVAVFAENAQSER 240

Query: 241 EQMTFKSLLKVSEIDVKNNNVVEVGNIFD 270
           E+M +KSLL+VSEIDVKNN V VE GNIFD
Sbjct: 241 EQMTFKSLLKVSEIDVKNNNVVVEVGNIFD 270
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1833

A DNA sequence (GBSx1940) was identified in *S.agalactiae* <SEQ ID 5689> which encodes the amino acid sequence <SEQ ID 5690>. This protein is predicted to be cmp-binding-factor 1. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1753(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC44803 GB:U21636 cmp-binding-factor 1 [Staphylococcus aureus]
 Identities = 140/310 (45%), Positives = 195/310 (62%), Gaps = 6/310 (1%)

Query: 3   INQMKKDELFEGFYLIKKAEVRKTRAGKDFIAFTFQDDTGEIGSNMWDAQTYNVEEFVAG 62
           I  +   +  + F+L+ KA    T  GKD++    QD +GEI    W A   ++
Sbjct: 4   IENLNPGDSVDHFFLVHKATQGVTAQGKDYMTLHLQDKSGEIEAKFWTATKNDMATIKPE 63

Query: 63  KIVHMKGRREVYNGTPQ--VNQITLRNIKDGEPNDPRDFKEKPPINVDNVREYMEQMLFK 120
           +IVH+KG    Y G  Q  VNQI L   +D      +   F  +  P++    ++E +   L
Sbjct: 64  EIVHVSGDIINYRGNKQMKVNQIRLATTEDQLKTE--QFVDGAPLSPAEIQESISHYLLD 121

Query: 121 IENATWQRVVRALYRKYNKEFFTYPAAKTNHHAFSSGLAYHTATMVRLADSIGDIYPELN 180
           IENA  QR+ R L +KY + F+TYPAA ++HH F SGL+YH  TM+R+A SI DIYP LN
Sbjct: 122 IENANLQRITRHLLKKYQERFYTYPAASSHHHNFASGLSYHVLTMLRIAKSICDIYPLLN 181

Query: 181 KSLMFAGIMLHDLAKVIELSGPDNTEYTIRGNLIGHISLIDELTKILAELNIDDTKEEV 240
           KSL+++GI+LHD+ KV ELSGP  T YT+ GNL+GHIS+  +E+ +   ELNI+   EE+
Sbjct: 182 KSLLYSGIILHDIGKVRELSGPVATSYTVEGNLLGHISIADEVVEAARELNIEG--EEI 239

Query: 241 TVLRHVILSHNGQLEYGSPVRPRIMEAEIIHMIDNIDANMMMMTALNRVNEGEMTNRIF 300
            +LRH+ILSHNG+LEYGSP  P + EAEI+  IDNIDA M M   A  + ++G+ T++IF
Sbjct: 240 MLLRHMILSHNGKLEYGSPKLPYLKEAEILCYIDNIDARMNMFEKAYKKTDKSQPTDKIF 299

Query: 301 AMDNRSFYKP 310
            ++NR FY P
Sbjct: 300 GLENRRFYNP 309
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5691> which encodes the amino acid sequence <SEQ ID 5692>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1822(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 275/311 (88%), Positives = 300/311 (96%)

Query: 1   MKINQMKKDELFEGFYLIKKAEVRKTRAGKDFIAFTFQDDTGEIGSNMWDAQTYNVEEFV 60
           MKINQMKKD+LFEGFYLIK AEVRKTRAGKDFI+ TFQDDTGEIGSN+WDAQ YNVEEF
Sbjct: 1   MKINQMKKDQLFEGFYLIKSAEVRKTRAGKDFISLTFQDDTGEIGSNLWDAQPYNVEEFT 60

Query: 61  AGKIVHMKGRREVYNGTPQVNQITLRNIKDGEPNDPRDFKEKPPINVDNVREYMEQMLFK 120
           AGK+V MKGRREVYNGTPQVNQITLRN++ GEPNDP+DFKEK P++V  VR+Y+EQMLFK
Sbjct: 61  AGKVVFMKGRREVYNGTPQVNQITLRNVRPGEPNDPKDFKEKAPVSVTEVRDYLEQMLFK 120
```

```
Query: 121 IENATWQRVVRALYRKYNKEPFTYPAAKTNHHAPESGLAYHTATMVRLADSIGDIYPELN 180
           IENATWQR+VRALYRKY+KEP+TYPAAKTNHHAPESGLAYHTATMVRLADSIGDIYP+LN
Sbjct: 121 IENATWQRIVRALYRKYDKEPFTYPAAKTNHHAPESGLAYHTATMVRLADSIGDIYPDLN 180

Query: 181 KSLMFAGIMLHDLAKVIELSGPDNTEYTIRGNLIGHISLIDEELTKILAELNIDDTKEEV 240
           KSL+FAGIMLHDLAKVIEL+GPDNTEYT+RGNLIGHISLI+EE+TK++EL IDDTKEEV
Sbjct: 181 KSLLFAGIMLHDLAKVIELTGPDNTEYTVRGNLIGHISLINEEITKVISELQIDDTKEEV 240

Query: 241 TVLRHVILSHHGQLEYGSPVRPRIMEAEIIHMIDNIDAMMMMTTALNRVNEGEMTNRIF 300
             VLRHVILSHHGQLEYGSPVRPRIMEAEIIHMIDNIDAMMMMTTAL+RV+EGEMTNRIF
Sbjct: 241 IVLRHVILSHHGQLEYGSPVRPRIMEAEIIHMIDNIDAMMMMTTALSRVSEGEMTNRIF 300

Query: 301 AMDNRSFYKPN 311
           AMDNRSFYKPN
Sbjct: 301 AMDNRSFYKPN 311
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1834

A DNA sequence (GBSx1941) was identified in *S.agalactiae* <SEQ ID 5693> which encodes the amino acid sequence <SEQ ID 5694>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.59   Transmembrane    2 -  18 (   1 -  22)

----- Final Results -----
                bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5695> which encodes the amino acid sequence <SEQ ID 5696>. Analysis of this protein sequence reveals the following:

```
      Possible site: 17

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.05   Transmembrane    3 -  19 (   1 -  26)

----- Final Results -----
                bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 309/424 (72%), Positives = 370/424 (86%), Gaps = 3/424 (0%)

Query: 1   MLVILIIVLASLTVTISYQKMTELTKSVEKQLEDNADNLSDQLTYQIEVAQKDQILLTL 60
           +++ +L++VL  L   ++    K+  L   + + LE NADNLSDQ+TYQ++ A K Q+L L
Sbjct: 3   LILFLLVLVLLGLGAYLLF--KVNGLQHQLAQTLEGNADNLSDQMTYQLDTANKQQLLEL 60

Query: 61  TNQLKNMQQRIYQLLTDMRTELNQHLTESRDRSDSDKRLELINSNLSQSVQKMQDSNEKRLD 120
           T  +NR Q  +YQ LTD+R  L++ L++SRDRSDSDKRLE IN  ++QS++ MQ+SNEKRL+
Sbjct: 61  TQLMNRQQAGLYQQLTDIRDVLHRSLSDSRDRSDSDKRLEKINQQVNQSLKNMQRSNEKRLE 120

Query: 121 QMRQTVEEKLEKTLQTRLQTSFETVSRQLESVNQGLGEMKTVAQDVGTLNKVLSNTKTRG 180
           +MRQ VEEKLE+TL+ RL  SF++VS+QLESVN+GLGEM++VAQDVGTLNKVLSNTKTRG
Sbjct: 121 KMRQIVEEKLEETLKNRLHASFDSVSKQLESVNKGLGEMRSVAQDVGTLNKVLSNTKTRG 180

Query: 181 ILGELQLGQIIEDIMTVSQYEREFPTVSGSSERVEYAIKLPGNQGDVIYLPIDSKPPLE 240
           ILGELQLGQIIEDIMT SQYEREF TVSGSSERVEYAIKLPGNQGD YIYLPIDSKPPLE
Sbjct: 181 ILGELQLGQIIEDIMTSSQYEREFVTVSGSSERVEYAIKLPGNQGDGYIYLPIDSKPPLE 240
```

```
Query: 241 DYYRLEDAYELGDKVQIELYRKSLLASIRKFAKDINNKYLNPPETNFGIMPLPTEGLYS 300
           DYYRLEDAYE+GDK+ IE  RK+LLA+I++FAKDI+ KYLNPPETNFG+MPLPTEGLYS
Sbjct: 241 DYYRLEDAYEVGDKLAIEASRKALLAAIKRFAKDIRKKYLNPPETNFGVMFLPTEGLYS 300

Query: 301 EVVRNATFFDSLRRDENIVVAGPSTLSALLNSLSVGFKTLNIQKNANDISKILGNVKVEF 360
           EVVRNA+FFDSLRR+ENIVVAGPSTLSALLNSLSVGFKTLNIQKNA+DISKILGNVK+EF
Sbjct: 301 EVVRNASFFDSLRREENIVVAGPSTLSALLNSLSVGFKTLNIQKNADDISKILGNVKLEF 360

Query: 361 GKPGGMLSKAQKQLNTASKSIDSLLTVGTNALIEVLNTVEKHQDQATFSLLNLPTTEEEE 420
            KPGG+L+KAQKQ+NTA+ ++D L++TRTNAI+EVLNTVEKHQDQATFSLLNLPTTEEEE
Sbjct: 361 DKPGGLLAKAQKQMNTANNTLDQLISTRTNAIVEALNTVETYQDQATKSLLNNPLLEEEN 420

Query: 421 INEN 424
             NEN
Sbjct: 421 -NEN 423
```

SEQ ID 5694 (GBS88) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 2; MW 48kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1835

A DNA sequence (GBSx1942) was identified in *S.agalactiae* <SEQ ID 5697> which encodes the amino acid sequence <SEQ ID 5698>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2722(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13453 GB:Z99112 yloS [Bacillus subtilis]
 Identities = 75/217 (34%), Positives = 109/217 (49%), Gaps = 12/217 (5%)

Query: 1   MTKIALFAGG------DLTYFEYDFDYFVGIDRGSLFLLKKGLSLDMAVGDFDSITEDEL 54
            M  I + AGG      DLT +  +  ++G+D+G++ LL  G+      A GDFDSITE E
Sbjct: 1   MKTINIVAGGPKNLIPDLTGYTDEFTLWIGVDKGTVTLLDAGIIPVEAFGDFDSITEQER 60

Query: 55  LYIKHYCSNIVSASAEKNDTDTELALKTIPKEPPEAQVTVPGAFGGRIDHMMSNIFLPSD 114
              I+      +    AEK+ TD +LAL     ++ P+  + +PG  GGR DH + NI L
Sbjct: 61  RRIEKAAPALHVYQAEKDQTDLDLALDWALEKQPDI-IQIPGITGGRADHFLGNIQLLYK 119

Query: 115 RDLEPPMSQIRLKDEQNIVTYLPSGKNQVSRIEGMSYVSPMPESES--TLQISGAKYELN 172
                    +IRL D+QN +    P G+  + + E    Y+SF+P SE     L ++G KY LN
Sbjct: 120 GVKTNI--KIRLIDKQNHIQMPPPGEYDIEKDENNKRYISFIPFSEDIHELTITGPKYPLN 177

Query: 173 KSNY-FKKKWYSSNKEWTSPIEVELKDGYLIIIYSKD 208
                +    + SNE + S        G LI+I S D
Sbjct: 178 NCHITLGSTLCISNELIHSRGTPSFAKGILIMIRSTD 214
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5699> which encodes the amino acid sequence <SEQ ID 5700>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
bacterial cytoplasm --- Certainty=0.2467(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 130/208 (62%), Positives = 166/208 (79%)

Query: 1   MTKIALFAGGDLTYFEYDFDYFVGIDRGSLFLLKNGLSLDMAVGDFDSITEDELLYIKHY 60
           M+K+ALFAGGDL+Y   DFDYFVGIDRGSLFLL+NGL L+MAVGDFDS+++      IK
Sbjct: 1   MSKVALFAGGDLSYISRDFDYFVGIDRGSLFLLENGLPLNMAVGDFDSVSQKAFTDIKEK 60

Query: 61  CSNIVGASQEKNDTDTPELALKTIPKEFPEAQVTVFGAPGGRIDHMMSNIFLPSDRDLEPF 120
             ++A  EKNDTDTELALK +F  FPEA+VT+FGAPGGR+DH++SNIFLPSD  + PF
Sbjct: 61  AELFITAHPEKNDTDTELALKEVFARFPEAEVTIFGAPGGRMDHLLSNIFLPSDPGIAPF 120

Query: 121 MSQIRLKDEQNIVTYLPSGKNQVSRIEGMSYVSPMPESESTLQISGAKYELNKSNYPKKK 180
           M+QI L+D+QN++TY P+G++ + + EGM+YV+PM E E+ L I+GAK+EL + N+PKKK
Sbjct: 121 MAQIALRDQQNMITYRPAGQHLIHQEEGMTYVAPMAEGEADLTITGAKFELTQDNFPKKK 180

Query: 181 MYSSNEFMTSPIEVELKDGYLIIIYSKD 208
           +YSSN F+  PI V L  GYLIII SKD
Sbjct: 181 IYSSNAFIHQPITVSLPSGYLIIIQSKD 208
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1836

A DNA sequence (GBSx1943) was identified in *S.agalactiae* <SEQ ID 5701> which encodes the amino acid sequence <SEQ ID 5702>. This protein is predicted to be ribulose-phosphate 3-epimerase (rpe). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.59   Transmembrane  124 - 140 ( 124 - 141)

----- Final Results -----
                bacterial membrane --- Certainty=0.1638(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06221 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 113/211 (53%), Positives = 153/211 (71%)

Query: 5   KIAPSILAADYANFANELKRIEETTAEYVHIDIMDGQFVPNISFGADVVSSMRKHSKLVF 64
           KIAPSIL+AD+AN  ME++ +E   A+Y+H+D+MDG FVPNI+ G  +V ++R  + L
Sbjct: 3   KIAPSILSADFANLGNEIQDVERGGADYIHVDVMDGHFVPNITIGPLVDAIRPVTTLPL 62

Query: 65  DCHLMVVDPERYIEAFAQAGADIMTIHVRATKHIHGALQKIKSAGMKAGVVINPGTPVES 124
           D HLM+  P+ YI AFA+AGADI+T+HVEA  H+H  L  IKE+G+KAGVV+NP TPV S
Sbjct: 63  DVHLMIEQPDGYIPAFAKAGADIITVHVEACPHLHRTLHLIKESGVKAGVVLNPATPVSS 122

Query: 125 LIPILDLVDQILIMTVNPGFGGQAFIPKMMSKVETVAAWRKEYGHHYDIEVDGGIDNTTI 184
           +  +L  VD +L MTVNPGFGGQ FIP ++ K+E +A+ +KE G  ++IEVDGG++  T
Sbjct: 123 IQHVLSDVDMVLFMTVNPGFGGQRFIPSVLPKLKELASLKKEQGLTFEIEVDGGVNEETA 182

Query: 185 KAAAEAGANVFVAGSYLFKASDLPAQVETLR 215
           K   EAGANV VAGS +F   D  A ++ +R
Sbjct: 183 KQCVEAGANVLVAGSAVFNEEDRAAAIKGIR 213
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5703> which encodes the amino acid sequence <SEQ ID 5704>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0072(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 183/219 (83%), Positives = 198/219 (89%)

Query: 1   MSTNKIAPSILAADYANFANELKRIEETTAEYVHIDIMDGQFVPNISFGADVVSSMRKHS 60
           MST KIAPSILAADYANFA+EL RIEET AEYVHIDIMDGQFVPNISFGADVV+SMRKHS
Sbjct: 1   MSTLKIAPSILAADYANFASELARIEETDAEYVHIDIMDGQFVPNISFGADVVASMRKHS 60

Query: 61  KLWFDCHLMVVDPERYIEAFAQAGADIMTIHVEATKHIHGALQKIKEAGMKAGVVINPGT 120
           KLWFDCHLMVVDPERY+EAFAQAGADIMTIH E+T+HIHGALQKIK AGMKAGVVINPGT
Sbjct: 61  KLVFDCHLMVVDPERYVEAFAQAGADIMTIHTESTRHIHGALQKIKAAGMKAGVVINPGT 120

Query: 121 PVESLIPILDLVDQVLLIMTVNPGFGGQAFIPEMMSKVKTVAAWRKEYGHHYDIEVDGGID 180
           P  +L P+LDLVDQ+LIMTVNPGFGGQAFIPE + KV TVA WR E G  +DIEVDGG+D
Sbjct: 121 PATALEPLLDLVDQVLIMTVNPGFGGQAFIPECLEKVATVAKWRDEKGLSFDIEVDGGVD 180

Query: 181 NTTIKAAAEAGANVFVAGSYLFKASDLPAQVETLRVALD 219
           N TI+A  EAGANVFVAGSYLFKASDL +QV+TLR AL+
Sbjct: 181 NKTIRACYEAGANVFVAGSYLFKASDLVSQVQTLRTALN 219
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1837

A DNA sequence (GBSx1944) was identified in *S.agalactiae* <SEQ ID 5705> which encodes the amino acid sequence <SEQ ID 5706>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2098(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13451 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 148/296 (50%), Positives = 202/296 (68%), Gaps = 14/296 (4%)

Query: 2   QGRIVKSLAGFYYV----ESDGVVYQTRARGNPRKKGQIPYVGDWVEFSSQDQSEGYILS 57
           +G+I+K+L+GFYYV    E    V Q R RG PRK     P VGD+V + +++  EGY++
Sbjct: 3   EGKIIKALSGFYYVLDESEDSDKVIQCRGRGIPRKNKITPLVGDYVVYQAENDKRGYLME 62

Query: 58  IEERKNSLVRPPIVNIDQAVVIMGAKEPDFNANLLDRFLVLLEYKMIQPIIYISKLDLLD 117
           I+ER N L+RPPI N+DQAV++ SA +P F+  LLDRFLVL+E   IQPII I+K+DL++
Sbjct: 63  IKERTNELIRPPICNVDQAVLVFSAVQPSFSTALLDRFLVLVEANDIQPIICITKMDLIE 122

Query: 118 DLVVIDDIR---EHYQNIGY-VFCYSQEE------LLPLLANKVTVFMGQTGVGKSTLLN 167
           D    D I+   E Y+NIGY V+  S ++       ++P   +K TVF GQ+GVGKS+LLN
Sbjct: 123 DQDTEDTIQAVASDYRNIGYDVVLTSSKDQDSLADIIPHPQDKTTVFAGQSGVGKSSLLN 182

Query: 168 KIAPELKLETGRISGSLGRGRHTTRAVSFYNVHKGKIALTPGFSSLDYEVDNAEDLNESF 227
```

```
              I+PEL L T EIS  LGRG+HTTR V   +   G +ÄDTPGFSSL++      E+L  +F
Sbjct: 183 AISPELGLRTNEISEHLGRGKHTTRHVELIHTSGGLVADTPGFSSLEFTDIEERELGYTF 242

Query: 228 PELRRLSHPCKFRSCTHTHRPKCAVKEALTQGQLWQVRYDNYLQFLSEIESRRETY 283
           P++R  S  CKFR C H  EPKCAVK+A+  G+L Q RYD+Y++F++EI+ R+  Y
Sbjct: 243 PDIREKSSSCKFRGCLHLKEPKCAVKQAVEDGKLKQYRYDHYVEPMTEIKDRKPRY 298
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5707> which encodes the amino acid sequence <SEQ ID 5708>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2290(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 230/290 (79%), Positives = 257/290 (88%)

Query: 1   MQGRIVKSLAGFYYVESDGVVYQTRARGNFRKKGQIPYVGDWVEFSSQDQSEGYILSIEE 60
           +QG+I+KSLAGFYYVES+G VYQTRARGNFRK+G+  PYVGD V+FS++D SEGYILS+I
Sbjct: 1   LQGKIIKSLAGFYYVESEGQVYQTRARGNFRKRGETPYVGDIVDFSAEDNSEGYILAIHP 60

Query: 61  RKNSLVRPPIVNIDQAVVIMSAKEPDFNAALLDRFLVLLEYKMIQPIIYISKLDLLDDLV 120
           RKNSLVRPPIVNIDQAVVIMSAKEP+FN+NLLDRFL+LLE+K I  P++YISK+DLLD
Sbjct: 61  RKNSLVRPPIVNIDQAVVIMSAKEPEFNSNLLDRFLILLEHKAIHPVVYISKMDLLDSPE 120

Query: 121 VIDDIREHYQNIGYVFCYSQEELLPLLANKVTVFMGQTGVGKSTLLNKIAPELKLETGEI 180
            I  I   YQ IGY F  S EELLPLLA+K+TVFMGQTGVGKSTLLN+IAPEL LE GEI
Sbjct: 121 EIKAIGRQYQAIGYDFVTSLEELLPLLADKITVFMGQTGVGKSTLLNRIAPELALEIGEI 180

Query: 181 SGSLGRGRHTTRAVSFYNVHKGKIADTPGFSSLDYEVDNAEDLNESFPELRRLSHFCKFR 240
           S SLGRGRHTTRAVSFYN H GKIADTPGFSSLDY++ NAEDLNE+FPELRRLSH CKFR
Sbjct: 181 SDSLGRGRHTTRAVSFYNTHGGKIADTPGFSSLDTDIANAEDLNRAFPELRRLSHSCKFR 240

Query: 241 SCTHTHRPKCAVKEALTQGQLWQVRYDNYLQFLSEIESRRETYKKVIKRK 290
           SCTHTHEPKCAVK AL  G+LW VRY++YLQFLSEIE+RRETYKKVIKRK
Sbjct: 241 SCTHTHEPKCAVKAALSTGELWPVRYEHYLQFLSEIENRRETYKKVIKRK 290
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1838

A DNA sequence (GBSx1945) was identified in *S.agalactiae* <SEQ ID 5709> which encodes the amino acid sequence <SEQ ID 5710>. This protein is predicted to be rRNA. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.43   Transmembrane  259 - 275 ( 259 - 275)

----- Final Results -----
                bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15937 GB:Z99124 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 95/278 (34%), Positives = 147/278 (52%), Gaps = 16/278 (5%)
```

```
Query: 14  SYFACPKCQNPLIKESN-SLKCSDN-HCPDLSKPGYVNLLGGKKVDEHYDKKSFENR-QL 70
           S F CP C + +   S  SL C++  H FDLS+ GYVN L  K V   Y  + FE R +L
Sbjct: 8   SMFRCPLCDSSMDAASGKSLICTERGHTFDLSRHGYVNFLT-KPVKTSYGAELFEARSRL 66

Query: 71  VLENGYYNHILEAISKVLENNSQFH---SVLDIGCGEGPYSRQLVNKHEKTFLAP----D 123
           + E G+++ + +AI++++ +     H   ++LD GCGEG +   L          A     D
Sbjct: 67  IGEGGFFDPLHDATAELISHPKSGHKAFTILDSGCGRGSHLNALCGFDYAGKAAIGTGID 126

Query: 124 ISKDSIQLAAKSDQSRLVKWFVSDLANLPIQDSSIDIILDIFSPANYKEFRRVLSDDGIL 183
           +SKD I  A+K+ +  W V+D+A  P  D   D++L IFSP+NY EF R+L +DG+L
Sbjct: 127 LSKDGILKASKAPKDLM--WAVADVARAPFHDRQFDVVLSIFSPSNYAEFHRLLKNDGML 184

Query: 184 VKVVPVAEHVQELREKASQYLKQKDYSNQKILDHPRENPEIISEQKVVQSYNCSQQERQA 243
           +KVVP ++++ ELR+       ++ YSN  ++ F  N       ++       QQ
Sbjct: 185 IKVVPRSDYLIELRQFLYTDSPRRTYSNTAAVERFTANAAHSRPVRLRYVKTLDQQAIHW 244

Query: 244 FIDMTPLLFSVDKTTIDW---ASISEITVGALIVIGKK 278
           + MTPL +S  K  +         ++ITV   I+IG K
Sbjct: 245 LLKMTPLAWSAPKDRVSLLKEMKSADITVDVDILIGMK 282
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1839

A DNA sequence (GBSx1946) was identified in *S.agalactiae* <SEQ ID 5711> which encodes the amino acid sequence <SEQ ID 5712>. This protein is predicted to be dimethyladenosine transferase (ksgA). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3257(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11818 GB:Z99104 dimethyladenosine transferase [Bacillus subtilis]
 Identities = 157/284 (55%), Positives = 215/284 (75%), Gaps = 2/284 (0%)

Query: 3   IADKTVTRAILERHGFTFKKSPQGNFLTVTNILQKIVDTARIDKGVNVIEIGPGIGALTE 62
           IA     T+ IL+++GF+FKKS GQNFL DTNIL +IVD AE+ +   VIEIGPGIGALTE
Sbjct: 5   IATPIRTKEILKKYGPSFKKSLGQNFLIDTNILNRIVDHAEVTEKTGVIEIGPGIGALTE 64

Query: 63  FLAENAAEVMAFEIDDRLIPILADTLARFDNVQVVNQDILKADLQTQIQA-FKNPDLPIK 121
            LA+ A +V+AFEID RL+PIL DTL+ ++NV V++QD+LKAD+++ I+  F++ D  I
Sbjct: 65  QLAKRAKKVVAFEIDQRLLPILKDTLSPYENVTVIHQDVLKADVKSVIEEQFQDCD-EIM 123

Query: 122 VVANLPYYITTPILMHLIESKIPFAEFVVMIQKEVADRISAMPNTKAYGSLSIAVQYYMT 181
           VVANLPYY+TTPI+M L+E  +P    VVN+QKEVA+R++A P++K YGSLSIAVQ+Y
Sbjct: 124 VVANLPYYVTTPIMKLLEEHLPLKGIVVMLQKEVASRMAADPSSKEYGSLSIAVQFYTE 183

Query: 182 AKVSPIVPRTVFVPAPNVDSAILKMVRRDQPVVSVQDRDFFRKVSKVAFVHRRKTLKNNL 241
           AK   IVP+TVFVP PNVDSA+++++ RD P V V++E FFF++ K +F  RRKTL NNL
Sbjct: 184 AKTVNIVPKTVFVPQPNVDSAVIRLILRDGPAVDVENESSFFFQLIKASFAQRRKTLLNNL 243

Query: 242 TSHPGKSEDTKAKLEKALEIAKIKPSIRGEALSIPDFASLADAL 285
           ++  + +  K+ +E+ LE   I    RGE+LSI +FA+L++ L
Sbjct: 244 VNNLPEGKAQKSTIEQVLEETNIDGKRRGESLSIEEFAALSNGL 287
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5713> which encodes the amino acid sequence <SEQ ID 5714>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2420(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 257/290 (88%), Positives = 275/290 (94%)

Query: 1   MRIADKTVTRALLKRHGFTFKKSFGQNFLTDTINILQKIVDTAEIDKGVNVIEIGPGIGAL 60
           MRIAD +VT+A+L+RHGFTFKKSFGQNFLTDTINILQKIVDTAEID+ VNVIEIGPGIGAL
Sbjct: 9   MRIADYSVTKAVLDRHGFTFKKSFGQNFLTDTINILQKIVDTAEIDQNVNVIEIGPGIGAL 68

Query: 61  TEFLAENAAEVMAFEIDDRLIPILADTLARFDNVQVVNQDILKADLQTQIQAFKNPDLPI 120
           TEFLAENAAEVMAFEIDDRL+PILADTL  FDNVQVVNQDILKADLQTQI+ FKNPDLPI
Sbjct: 69  TEFLAENAAEVMAFEIDDRLVPILADTLRDFDNVQVVNQDILKADLQTQIKQFKNPDLPI 128

Query: 121 KVVANLPYYITTPILMHLIESKIPFAEFVVMIQKEVADRISAMPNTKAYGSLSIAVQYYN 180
           KVVANLPYYITTPILMHLIESKIPF EFVVM+Q+EVADRISA PNTKAYGSLSIAVQYYM
Sbjct: 129 KVVANLPYYITTPILMHLIESKIPFQEFVVMMQREVADRISAEPNTKAYGSLSIAVQYYM 188

Query: 181 TAKVSFIVPRTVFVPAPNVDSAILKMVRRDQPVVSVQDEDFFFRVSKVAFVHRRKTLWNN 240
           TAKV+FIVPRTVFVPAPNVDSAILKMVRRDQP++ V+DEDFFFRVS+++FVHRRKTLWNN
Sbjct: 189 TAKVAFIVPRTVFVPAPNVDSAILKMVRRDQPLIEVKDEDFFFRVSRLSFVHRRKTLWNN 248

Query: 241 LTSHFGKSEDTKAKLEKALEIAKIKPSIRGEALSIPDFASLADALKEVGI 290
           LTSHFGKSED KAKLEK L +A IKPSIRGEALSI DF  LADALKEVG+
Sbjct: 249 LTSHFGKSEDIKAKLEKGLALADIKPSIRGEALSIQDFGKLADALKEVGL 298
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1840

A DNA sequence (GBSx1947) was identified in *S.agalactiae* <SEQ ID 5715> which encodes the amino acid sequence <SEQ ID 5716>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0736(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1841

A DNA sequence (GBSx1948) was identified in *S.agalactiae* <SEQ ID 5717> which encodes the amino acid sequence <SEQ ID 5718>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.3031(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11817 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 81/179 (45%), Positives = 117/179 (65%), Gaps = 4/179 (2%)

Query: 7   IQEVIVVEGKDDTANLRRFYNVDTYETRGSAIDEDDLERIERLHNLRGVIVFTDFDVNGE 66
            I+E+IVVEG+DDTA ++    + DT ET GSAID+  +++I       RGVI+ TDFD+ GE
Sbjct: 3   IKEIIVVEGRDDTARIKLAVDADTIETNGSAIDDHVIDQIRLAQKTRGVIILTDFDPEGE 62

Query: 67  RIRKIIMNAIPTVRHAFLNRDEAKPGSKTKGRSLGVEHASFEDLQKALSKVTQHFDDEDH 126
            +IRK I  A+P  +HAFL +  AKP +K   R +GVEHAS E ++  L  V +  + +
Sbjct: 63  KIRKTISEAVPGCKHAFLPKHLAKPKNK---RGIGVEHASVESIRACLENVHEEMEAQPS 119

Query: 127 FDITQADLIRWGFITASDSRKRRSYLGNQLRIGYSNGKQLLKRLRLFGVTKAEVEECME 185
             DI+  DLI  G I    ++ RRE LG+ L+IGY+NGKQL KRL++F + K++    ++
Sbjct: 120 -DISAEDLIHAGLIGGPAAKCRRERLGDLLKIGYTNGKQLQKRLQMFQIKKSDFMSALD 177
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5719> which encodes the amino acid sequence <SEQ ID 5720>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.1474(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 146/187 (78%), Positives = 165/187 (88%)

Query: 1   MMKKIDIQEVIVVEGKDDTANLRRFYNVDTYETRGSAIDEDDLERIERLHNLRGVIVFTD 60
            + +KI+IQEV+VVEGKDDTANLRRFY VDTYETRGSAI E+DLERI RL++LRGVIV TD
Sbjct: 15  LTEKINIQEVLVVEGKDDTANLRRFYEVDTYETRGSAITEEDLERINRLNDLRGVIVLTD 74

Query: 61  PDYNGERIRKIIMNAIPTVRHAFLNRDEAKPGSKTKGRSLGVEHASFEDLQKALSKVTQH 120
            PDYNGERIRK+IM A+PT RHAFLNR+EA P SK+KGRSLGVEHA+FEDLQKAL+ VTQ
Sbjct: 75  PDYNGERIRKLIMAAVPTARHAFLNRNEAVPSSKSKGRSLGVEHANFEDLQKALAHVTQQ 134

Query: 121 FDDEDHFDITQADLIRWGFITASDSRKRREYLGNQLRIGYSNGKQLLKRLRLFGVTKAEV 180
            +DDE +FDI Q DLIR G + ASDSRKRREYLG +LRIGY+NGKQLLKRL LFG+T AEV
Sbjct: 135 YDDESYFDIRQTDLIKLGLIMASDSRKRREYLGSKLRIGYANGKQLLKRLELFGITLAEV 194

Query: 181 EECMEGY 187
            EE ME Y
Sbjct: 195 EEVMETY 201
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1842

A DNA sequence (GBSx1949) was identified in *S.agalactiae* <SEQ ID 5721> which encodes the amino acid sequence <SEQ ID 5722>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4955(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10139> which encodes amino acid sequence <SEQ ID 10140> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11815 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 115/254 (45%), Positives = 172/254 (67%)

Query: 28  IFDTHTHLNVENFEGKIDEEINLASRLGVTKMNVVGPDQDTISKSLELSSQYAQVYSTIG 87
           +FDTH HLN E ++  +++E I  A    V ++ VVGFD+ TI++++E+  +Y  +Y+ IG
Sbjct: 2   LFDTHAHLNAEQYDTDLEEVIERAKAEKVSRIVVVGFDRPTITRAMEMIEBYDFIYAAIG 61

Query: 88  WHPTEAGSYDDNIESMIISHLENPKVIALGEIGLDYYWMEDPKDIQIEVFKRQIELSKEY 147
           WHP +A    +  I     + KV+A+GE+GLDY+W D K PKDIQ EVF+ QI L+KE
Sbjct: 62  WHPVDADMTEEDLAWIKELSAHEKVVAIGEMGLDYHWDKSPKDIQKEVFNQIALAKEV 121

Query: 148 NLPFVVHTRDALEDTYEVIKESGVGPFGIMHSFSGSLEMAQKFIDLGMMISFSGVVTFK 207
           NLP ++H RDA ED   ++KE G    GGIMH F+GS E+A++ + +   +SF G VTFK
Sbjct: 122 NLPIIIHNRDAYEDVVTILKSEGARAVGGIMHCFTGSAEVARECMKMNFYLGFGGPVTFK 181

Query: 208 KALDVQEAARELPLDKILVSTDAPYLAPVPKRGRENKTAYTRYVVSKIAELRGITVSEVA 267
            A   +E  +E+P D++L+ETD P+L P P RG+ N+ +Y +YV E+IAEL+ +T EE+A
Sbjct: 182 NAKKPKEVVKEIPNDRLLIETDCPFLTPHPFRGKRNEPSYVKYVAEQIAELKEMTFEEIA 241

Query: 268 EATYQNAVRIFRLD 281
           T +NA R+FR++
Sbjct: 242 SITTENAKRLFRIN 255
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5723> which encodes the amino acid sequence <SEQ ID 5724>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2817(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 190/258 (73%), Positives = 227/258 (87%)

Query: 24  DMIKIFDTHTHLNVENFEGKIDEEINLASELGVTKMNVVGPDQDTISKSLELSSQYAQVY 83
           + + IFDTHTHLNV  F+G   ER+ LA E+GV   NVVGPDQ TIS +L L+++YA +Y
Sbjct: 38  EKLTIFDTHTHLNVAEFQGHIEEELTLAQEMGVAYHNVVGPDQATISGALTLANKYANIY 97

Query: 84  STIGWHPTEAGSYDDNIESMIISHLENPKVIALGEIGLDYYWMEDPKDIQIEVFKRQIEL 143
           +TIGWHPTEAGSY + +E  I+S L + KVIALGEIGLDYYWMEDPK++QIEVFKRQ++L
Sbjct: 98  ATIGWHPTEAGSYSEAVESAIVSQLSHSKVIALGEIGLDYYWMEDPKEVQIEVFKRQMQL 157

Query: 144 SKEYNLPFVVHTRDALEDTYEVIKESGVGPFGIMHSFSGSLEMAQKFIDLGMMISFSGV 203
           +K++PFVVHTRDALEDTYEVIK +GVGP GGIMHS+SGSLEMA++FI+LGMMISFSGV
Sbjct: 158 AKDNDLPFVVHTRDALEDTYEVIKAAGVGPKGIIMHSYSGSLEMARFFIELGMMISFSGV 217

Query: 204 VTFKKALDVQEAARELPLDKILVETDAPYLAPVPKRGRENKTAYTRYVVEKIAELRGITV 263
           VTFKKALD+QEAA+ LPLDKILVETDAPYL PVPKRG++N TAYTRYVV+KIAELRG+TV
Sbjct: 218 VTFKKALDIQEAAQHLPLDKILVETDAPYLTPVPKRGKQNHTAYTRYVVDKIAELRGMTV 277
```

```
Query: 264 EEVAEATYQNAVRIFKLD 281
          EEVA+AT  NA R+F+LD
Sbjct: 278 EEVAKATIANAKRVFKLD 295
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1843

A DNA sequence (GBSx1950) was identified in *S.agalactiae* <SEQ ID 5725> which encodes the amino acid sequence <SEQ ID 5726>. This protein is predicted to be endosome-associated protein. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.5142(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1844

A DNA sequence (GBSx1951) was identified in *S.agalactiae* <SEQ ID 5727> which encodes the amino acid sequence <SEQ ID 5728>. This protein is predicted to be CG17785 gene product. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4730(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1845

A DNA sequence (GBSx1952) was identified in *S.agalactiae* <SEQ ID 5729> which encodes the amino acid sequence <SEQ ID 5730>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4032(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB01041 GB:AB022220 gene_id:MLN21.14~unknown protein
          [Arabidopsis thaliana]
 Identities = 49/185 (26%), Positives = 85/185 (45%), Gaps = 46/185 (24%)

Query: 5   LTDLDRVNIAKQEVELGSQLDTLVKIMSQDKVLPIGKVAHVQ------DGGKETGEQIYT 58
           L  +D V+  + + ELGS+     + +M+         K+  V+       D  K+   Q++
Sbjct: 154 LEGIDSVDSGRVKIELGSRGLMDLCVMASKLAYENAKMNLVEFLDCWNDYQKQMSTQVFV 213

Query: 59  ITPNGTLDKPEDVKEVTVLPEGSTAPFGSDDWKTD----WVKNDIPIASKL---LLKKPG 111
            T     DK +D    + + P+G T PF  DDW TD     W+  ++P    KL    L+  G
Sbjct: 214 FT-----DKQKDANLIVISPRG-TEPFDADDWGTDFDYSWY--EVPNVGKLHMGFLEAMG 265

Query: 112 ----------------SQSVSHKQSTKQ-----LEQSAH-----LLKEVMNKYPNAKISVY 146
                          Q+ S ++ +K+      +E+SA+      +LK +++++ NA+  V
Sbjct: 266 LGNRDDTTTFHYNLFEQTSSEEENSKKNLLDMVERSAYYAVRVILKRLLSEHENARFVVT 325

Query: 147 GHSLG 151
           GHSLG
Sbjct: 326 GHSLG 330
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1846

A DNA sequence (GBSx1953) was identified in *S.agalactiae* <SEQ ID 5731> which encodes the amino acid sequence <SEQ ID 5732>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.97   Transmembrane   12 -  28 (   5 -  33)

----- Final Results -----
                bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10141> which encodes amino acid sequence <SEQ ID 10142> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8909> and protein <SEQ ID 8910> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:    14.01
GvH: Signal Score (-7.5): -5.55
     Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -8.97 threshold:  0.0
   INTEGRAL    Likelihood = -8.97   Transmembrane    6 -  22 (   1 -  27)
```

```
   PERIPHERAL  Likelihood =  9.49       84
 modified ALOM score:   2.29

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 8910 (GBS32) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 10 (lane 2; MW 15.6kDa).

GBS32-His was purified as shown in Figure 191, lane 8.

Example 1847

A DNA sequence (GBSx1954) was identified in *S.agalactiae* <SEQ ID 5733> which encodes the amino acid sequence <SEQ ID 5734>. This protein is predicted to be extramembranal protein (dltD). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.24   Transmembrane   12 -  28 (   4 -  31)

----- Final Results -----
                bacterial membrane --- Certainty=0.5097(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC29041 GB:AF050517 unknown [Streptococcus mutans]
 Identities = 242/421 (57%), Positives = 309/421 (72%), Gaps = 1/421 (0%)

Query: 1   MLKRLGKVFGPLVCALLLLVGLYFVFPVSQ-PHHLGKEKNSAVALTKAGPKSRVQKVRAF 59
           MLKRL  + GP+ CAL+L+  L   +P      H+  +EKN AVAL+ + FKS  +K+RA
Sbjct: 1   MLKRLWLILGPVFCALVLVFSLIMFYPAKHLSHNYNEEKNDAVALSPSSFKSTNKKMRAL 60

Query: 60  SDPKANFVPFFGSSEWLRFDAMHPSVLAEAYMRSYIPYLLGQKGAASLTQYYGIQQIKGQ 119
           SD +  FVPFFGSSEW R D MHPSVLAE YMRSY PYLLGQKG+ SL+ Y+G+QQI  Q
Sbjct: 61  SDKRHLFVPFFGSSEWQRIDNMHPSVLAERYMRSYRPYLLGQKGSTSLSHYFGMQQIGNQ 120

Query: 120 IKNKKAIYVISPQWFVRKGANKGAFQNYFSNDQTIRFLQNQTGTTYDRYAARRLLKLYPE 179
           IKNKKA+YVISPQWFV KG +  APQ YFS+=Q   FL NQTG+T DRYAA+RLL + P
Sbjct: 121 IKNKKAVYVISPQWFVPKGTSPIAFQQYFSSEQLADFLLNQTGSTADRYAAKRLLDIKPS 180

Query: 180 AEMSDLIEKVADQQKLSNKDKQRLKFNDWVFEKTDAIFSYLPLGKTYNQAIMPFVGKLPK 239
           +++  +I+K+A G+ L++ D+  L+       +K DA+F  L     Y + ++PHV KLPK
Sbjct: 181 SNLQGMIKKIAAGKTLNSFDRASLRLIKSFLKKEDALFGSLTPSDNYERRVLPFVKKLPK 240

Query: 240 AFSYNHLSRIASQDAKVATRSNQFGIDDRFYQTRIKKHLKKLKGSQRHFNYTKSPEFNDL 299
            FSY  LS+IAS+D +  T++NQF I+D FY  RIK  LK+LKG Q+  +Y +SPE+NDL
Sbjct: 241 HFSYGTLSQIASKDSQRILTKTNQFSINDHFYNKRIKGQLKRLKGFQKQLSYLQSPEYNDL 300

Query: 300 QLVLNEFSKQNTDVLFVIPPVNKKWTDYTGLDQKMYQKSVEKIKHQLQSQGGFNHIADLSR 359
           QL L + +K  T V+FVIPPVN KW +YTGL Q MYQK+VEKIK+QLQSQGF++IADLS+
Sbjct: 301 QLALTQLAKSKTKVIFVIPPVNAKWVEYTGLSQDMYQKTVEKIKYQLQSQGFDNIADLSK 360

Query: 360 DGSKPYFMQDTIHLGWNGWLSLDKHINPFLTERNSKPNYHINNKFLKKSWAKVTQRPSDYK 420
           +G +PYFMQDTIHLGWNGWL  DK +NPFL+++  +P Y INN FL K WA VTG P  +K
Sbjct: 361 NGDQPYFMQDTIHLGWNGWLAPDKEVNPFLSKKQLQPAYKINNHFLSKGWATVTGNPFQFK 421
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5735> which encodes the amino acid sequence <SEQ ID 5736>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood =-13.06   Transmembrane    7 -  23 (   1 -  31)

----- Final Results -----
                bacterial membrane --- Certainty=0.6222(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 209/410 (50%), Positives = 278/410 (66%)

Query: 1   MLKRLGKVPGPLVCALLLLVGLYFVFPVSQPHHLGKEKNSAVALTKAGFKSRVQKVRAPS 60
           MLKRL  + GPL+ A +L+V   F FP    H + +EK +AVA+T + FK+ + K +A S
Sbjct: 1   MLKRLWLILGSPLLIAFVLVVITIFSFPTQLDHSIAQEKANAVAITDSSFKNGLIKRQALS 60

Query: 61  DPKANFVPFFGSSEWLRFDAMHPSVLAEAYNRSYIPYLLGQKGAASLTQYYGIQQIKGQI 120
           D    FVPFFGSSEW R D+MHPSVLAE Y RSY P+L+G++G+ASL+ YYGIQQI  ++
Sbjct: 61  DETCRFVPFFGSSEWSRMDSMHPSVLAERYKRSYRPFLIGKRGSASLSHYYGIQQITNEM 120

Query: 121 KNKKAIYVISPQWFVRKGANKGAPQNYFSNDQTIRFLQNQTGTTYDRYAARRLLKLYPEA 180
           + KKAI+V+SPQWF  +G N  A Q Y SN Q I FL          ++AA+RLL+L P
Sbjct: 121 QKKKAIFVVSPQWFTAQGINPSAVQMYLSNTQVIEFLLKARTDKESQFAAKRLLELNPGV 180

Query: 181 SMSDLIEKVADGQKLSNKDKQRLKFNDWVPEKTDAIPSYLPLGKTYNQAIMPHVGKLPKA 240
           S S+L++KV+ G+ LS  D+  LK     V  + +++PS+L     Y + I+P V  LPK
Sbjct: 181 SKSNLLKKVSNGKSLSRLDRAILKCQHQVALRESLPSFLGKSTNYEKRILPRVKGLPKV 240

Query: 241 FSYNHLSRIASQDAKVATRSNQPGIDDRFVQTRIKKHLKKLKGSQRHFNYTKSPEFNDLQ 300
           FSY  L+ +A++  ++AT +N+PGI + FV+ RI       K  Q +++Y  SPE+ND Q
Sbjct: 241 FSYKQLNALATKRGQLATTNNRPGIKNTFYRKRIAPKYNLYKNFQVNYSYLASPEYNDFQ 300

Query: 301 LVLNEFSKQNTDVLFVIPFVNKKWTDYTGLDQKMYQKSVEKIKHQLQSQGFNHIADLSRD 360
           L+L+EF+K+ TDVLFVI PVNK W DYTGL+Q  YQ +V KIK QL+SQGF+ IAD S+D
Sbjct: 301 LLLSEFAKRKTDVLFVITFVNKAWADYTGLNQDKYQAAVRKIKFQLKSQGFNHIADFSKD 360

Query: 361 GGKPYFMQDTIHLGWNGWLELDKHINPFLTEENSKPNVHINNKFLKKSWA 410
           GG+ YFMQDTIHLGWNGWL  DK + PFL +   PNY +N  F  K WA
Sbjct: 361 GGESYFMQDTIHLGWNGWIAPDKKVQPFLETKQFVPNYKMNPYFYSKIWA 410
```

A related GBS gene <SEQ ID 8911> and protein <SEQ ID 8912> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:     15.50
GvH: Signal Score (-7.5): -4.52
     Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value: -10.24 threshold:  0.0
   INTEGRAL    Likelihood =-10.24   Transmembrane   12 -  28 (   4 -  31)
   PERIPHERAL  Likelihood =  8.33      301
 modified ALOM score:   2.55

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.5097(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
57.5/76.3% over 420aa
                                                    Streptococcus mutans
  GP|3403204| unknown Insert characterized
```

```
ORF00336(301 - 1560 of 1860)
GP|3403204|gb|AAC29041.1||AF050517(1 - 421 of 421) unknown {Streptococcus mutans}
%Match = 41.0
%Identity = 57.5  %Similarity = 76.2
Matches = 242  Mismatches = 99  Conservative Sub.s = 79

33          63          93          123         153         183         213         243
FSGFLDLLWFPQPHNK**GVL*WILNQRY*QLLMTYLWRMFLL*WMKTVLTQEF*TAWVLLN*LLSWKATLILLFRLRNL

273         303         333         363                     420         450         480
VVMTGTQLIKLLLE*RSSAMLKRLGKVPGPLVCALLLLVGLYFVPVSQ-PHHLGKEKNSAVALTKAGFKSRVQKVRAPS
                        |||||  ::||: |||:|:  |   :|  :  |:  :||| ||||: : |||  :|:||:|
                       MLKRLWLILGPVFCALVLVFSLIMFYPAKHLSHNYNEEKNDAVALSPSSFKSTNKKMRALS
                                 10          20          30          40          50          60

510         540         570         600         630         660         690         720
DPKANFVPFFGSSEWLRFDAMHPSVLARAYNRSYIPYLLGQKGAASLTQYYGIQQIKGQIKNKKAIYVISPQWFVRKGAN
|  :  |||:|||||| | | |||||||| ||||| ||||||||: |||::|:|:|||  |||||||:|||||||| || :
DKRHLFVPFFGSSEWQRIDNMHPSVLAERYNRSYRPYLLGQKGSTSLSHYFGMQQIGNQIKNKKAVYVISPQWFVPKGTS
                    80          90          100         110         120         130         140

750         780         810         840         870         900         930         960
KGAPQNYFSNDQTIRFLQNQTGTTYDRYAARRLLKLYPEASMSDLIEKVADQKLSNKDKQRLKFNDWVFEKTDAIFSYL
  |||  ||||::|   || ||||:| |||||:||| : |  :::  :|:|:| |: |::  |:  |::    ::| ||:|  |
PIAPQQYFSSEQLADFLLNQTGSTADRYAAKRLLDIKFSSNLQGMIKKIAAGKTLNSFDRASLRLIKSFLKKSDALFGSL
                    160         170         180         190         200         210         220

990         1020        1050        1080        1110        1140        1170        1200
PLGKTYNQAIMPHVGKLPKAFSYNHLSRIASQDAKVATRSNQFGIDDRFYQTRIKKHLKKLKGSQRHFNYTKSPEFNDLQ
 :   | : ::||| |||| |||  ||:|||:| :  |::||| :|  || ||| :||:|||  |::::| :|||:||||
TFSDNYERRVLPHVKKLPKHFSYGTLSQIASKDSQRLTKTNQFEINDHFYNKRIKGQLKRLKGFQKQLSYLQSPEYNDLQ
                    240         250         260         270         280         290         300

1230        1260        1290        1320        1350        1380        1410        1440
LVLNEFSKQNTDVLFVIPPVNKKWTDVTGLDQKMYQKSVEKIKHQLQSQGFNHIADLSRDGGKPYFMQDTIHLGWNGWLE
| | :::|  | |:|||||| || :|||| | ||||:|||||:|||||||:::||||::|  :|||||||||||||||
LALTQLAKSKTKVIFVIPPVNAKWVKYTGLSQDMYQKTVEKIKYQLQSQGFDNIADLSKNGDQPYFMQDTIHLGWNGWLA
                    320         330         340         350         360         370         380

1470        1500        1530        1560        1590        1620        1650        1680
LDKHINPFLTEENSKPNYHINNKFLKKSWAKYTGRPSDYK*IVESDDL*H*SY*SSFLISLYLVILR*LIHVL*FFIYNR
:|| :||||:::  :| | ||| || | |: ||| |  :|
FDKEVNPFLSKKQLQPAYKINDHFLSKKWATYTGNPFQFK
                    400         410         420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1848

A DNA sequence (GBSx1955) was identified in *S.agalactiae* <SEQ ID 5737> which encodes the amino acid sequence <SEQ ID 5738>. This protein is predicted to be d-alanyl carrier protein (dltC). Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1061(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC05776 GB:AF051356 D-alanyl carrier protein [Streptococcus mutans]
```

```
 Identities = 65/79 (82%), Positives = 74/79 (93%)

Query: 1  MDIKSEVLAIIDDLFMEDVSSMMDEDLFDAGVLDSMGTVELIVELESHFNIDIPIAEFGR 60
          MDIKSEVL IID+LFMEDVS MMDEDLFDAGVLDSMGTVELIVELE+HF+I +P++EFGR
Sbjct: 1  MDIKSEVLKIIDELFMEDVSDMMDEDLFDAGVLDSMGTVELIVELENHFDITVPVSEFGR 60

Query: 61 NDWNIANKIVAGVTELCNA 79
          +DWNIANKI+ G+TEL NA
Sbjct: 61 DDWNIANKIIEGITELRNA 79
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5739> which encodes the amino acid sequence <SEQ ID 5740>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3976(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 57/79 (72%), Positives = 65/79 (82%)

Query: 1  MDIKSEVLAIIDDLFMEDVSSMMDEDLFDAGVLDSMGTVELIVELESHFNIDIPIAEFGR 60
          M I+  V+ + D LFMEDVS MMDEDLFDAGVLDS+GTVELIVELES FNI +PI+EFGR
Sbjct: 1  MSIEETVIELFDRLFMEDVSEMMDEDLFDAGVLDSLGTVELIVELESTFNIKVPISEFGR 60

Query: 61 NDWNIANKIVAGVTELCNA 79
          +DWNI  KIV GV EL +A
Sbjct: 61 DDWNIVTKIVQGVEELQHA 79
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1849

A DNA sequence (GBSx1956) was identified in *S.agalactiae* <SEQ ID 5741> which encodes the amino acid sequence <SEQ ID 5742>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.55   Transmembrane   93 - 109 (  91 - 117)
   INTEGRAL    Likelihood = -7.64   Transmembrane   21 -  37 (  19 -  39)
   INTEGRAL    Likelihood = -6.79   Transmembrane  390 - 406 ( 387 - 410)
   INTEGRAL    Likelihood = -5.20   Transmembrane   41 -  57 (  40 -  59)
   INTEGRAL    Likelihood = -2.07   Transmembrane  203 - 219 ( 200 - 221)
   INTEGRAL    Likelihood = -1.65   Transmembrane   65 -  81 (  65 -  81)
   INTEGRAL    Likelihood = -0.75   Transmembrane  125 - 141 ( 125 - 141)

----- Final Results -----
               bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5743> which encodes the amino acid sequence <SEQ ID 5744>. Analysis of this protein sequence reveals the following:

```
     Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.14   Transmembrane  387 - 403 ( 382 - 409)
   INTEGRAL    Likelihood = -9.66   Transmembrane   18 -  34 (  15 -  37)
   INTEGRAL    Likelihood = -5.95   Transmembrane   64 -  80 (  63 -  81)
```

```
  INTEGRAL     Likelihood = -5.63   Transmembrane    92 - 108 (  89 - 114)
  INTEGRAL     Likelihood = -1.97   Transmembrane    40 -  56 (  40 -  56)

 ----- Final Results -----
                bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC05775 GB:AF051356 integral membrane protein [Streptococcus mutans]
 Identities = 246/413 (59%), Positives = 319/413 (76%)

Query: 1   MMKFFSHIPYMEPYGNPIYVVLLLAPLVYIGIYGKQRLGTYETLVGLVFLLPMPGGDH 60
           M+ FF ++P++E YGNP YF Y+ILA LP+ IG+F +KR   YE  VSL+FI+ M  G+
Sbjct: 1   MIDFFKNLPHLEAYGNPQYFFYIILAVLPIFIGLFFKKRFPLYEAFVSLIFIVLMLTGEK 60

Query: 61  YCQDLVAFLFYLLWQIISVFAYQKYRENANSAGVFYLAIAMALPLIWVKVAPLTGPSSQT 120
             Q+ A  FY++WQI  V++Y+ YR++ ++  +FYL + M++ PL  VK+ P    + Q+
Sbjct: 61  SHQIFALFFYIIWQIPCVYSYKFYRKSRDNKWIFYLHVFMSILPLSLVKITPAIWTNQQS 120

Query: 121 LPSFLGISYLTFKSIGMIIEMRDGTLQEVRLPDFIRFMIFFPTFSGGPIDRFRHFQEDYH 180
            LF FLGISYLTF+S+GMI+EMRDG L      +FIRFM+F PTFSGGPIDRFR F +DY
Sbjct: 121 LFGFLGISYLTFRSVGMIMEMRDGVLTSFTFWEPIRFMLFMPTFSGGPIDRFRRFNDDYE 180

Query: 181 KLPERDDYFAMLNKAVMVLMLGFLYKHIISYCLGGILLPLLENKALMVGGYFNKETILVM 240
            K+P++D+   ML ++V Y+MLGF YK +++  LG ++LP L+  AL  GG+FN  T+ VM
Sbjct: 181 KIPDKDELDMLEQSVHYIMLGFFYKFVLAQILGTMILPGLKSMALQKGWFNWFTLGVM 240

Query: 241 YVYGLNLFFDFAGYSMFAIGISYLLGIRTPENFNMPFLSASLKDFWNRWHMSLSFWFRDY 300
            YVYGL+LFFDFAGYSMFAI IS  +GI++P NFN PF S  LK+FWNRWHMSLSFWFRD+
Sbjct: 241 YVYGLDLFFDFAGYSMFAIAISNFMGIKSPTNFNQPFKSQDLKEFWNRWHMSLSFWFRDF 300

Query: 301 VFMRLVHLLIKHKTFKNRNVTSGVAYLVNMLVMGFWHGLTWYYIAYGLFHGIGLIINDAW 360
            VFMRLV +L+K+K FKNRNVTS VAY+VNML+MGFWHG+TWYYI YGLFHG+GL++NDAW
Sbjct: 301 VFMRLVKVLVKNKVFKNRNVTSSVAYIVNMLIMGFWHGVTWYYITYGLFHGVGLVLNDAW 360

Query: 361 IRKKKEINRHRKKKGLSLPLQSRAFHVLCIVVTHVVMFSLLLPGSVLNDLWF 413
            +RKKK +N+ RK K LSPL ++        L IV+TF+VVM S L+PSGFLNDLWF
Sbjct: 361 LRKKKRLNKERKAKNLSPLPENGWTRALGIVITFNVVMLSPLIPSGFLNDLWF 413
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 240/416 (57%), Positives = 317/416 (75%), Gaps = 5/416 (1%)

Query: 5   FLEKLPHLDVYGNPQYFFYLILAVLPIYIGLFFKKRFALYEIIFSLSFIVMMLTGSTFNQ 64
           F   +P+++ YGNP YF YLILA LP+ IG+F +KR + YE + SL FI+ M  G  + Q
Sbjct: 4   FFSHIPYMEPYGNPIYFVVLILAFLPVIIGIFKQKRLSTYETLVSLVFILFMPGSDHYQQ 63

Query: 65  LKSLLAYVVGQSLLVFIYKAYRKRFNHTLVFYVTVCLSIFPLFLVKLIPAISEDGHQSLF 124
           L + L Y++ Q + VF Y+ YR+  N   VFY+ + +++FPL  VK+ P ++    Q+LF
Sbjct: 64  LVAFLFYLLWQIISVFAYQKYRENANSAGVFYLAIAMALFPLIWVKVAP-LTGPSSQTLF 122

Query: 125 GFLGISYLTFRAVAMIIEMRDGVLKEFTLWEFLRFLLFFPTFSSGPIDRFKRFNEDYINI 184
            FLGISYLTF+++ MIIEMRDG L+E  L +F+RF++FFPTFSSGPIDRF+ F EDY +
Sbjct: 123 SFLGISYLTFKSIGMIIEMRDGTLQEVRLPDFIRFMIFFPTFSSGPIDRFRHFQEDYHKL 182

Query: 185 PDRNELLDMLGQAIHYIMLGFLYKFILAYIFGSLIMPPLKELALEQGGVFNWPTLGVMYA 244
           P+R++   ML +A+ YLMLGFLYK I++Y  G +++P L+  AL  GG FN  T+ VMY
Sbjct: 183 PERDDYFAMLNKAVMYLMLGFLYKHIISYCLGGILLPLLENKALMVGGYFNKETILVMYV 242

Query: 245 PGFDLFFDFAGYTMFALAISNLMGIKSPINFDKPFKSRDLKEFWNRWHMSLSFWFRDFVF 304
           +G +LFFDFAGY+MFA+ IS L+GI++P NF+ PF S  LK+FWNRWHMSLSFWFRD+VF
Sbjct: 243 YGLNLFFDFAGYSMFAIGISYLLGIRTPENFNMPFLSASLKDFWNRWHMSLSFWFRDYVF 302

Query: 305 MRLVKLLVKNKVFKNRNVTSSVAYIINMLLMGFWHGLTWYYIAYGLFHGIGLVINDAMVR 364
           MRLV LL+K+K FKNRNVTS VAY++NML+MGFWHGLTWYYIAYGLFHGIGLI+NDAWIR
Sbjct: 303 MRLVHLLIKHKTFKNRNVTSGVAYLVNMLVMGFWHGLTWYYIAYGLFHGIGLIINDAWIR 362
```

```
Query: 365 KKKNINKERRLAKKPLLP--ENKWTYALGVFITFNVVMFSFLIFSGFLDLLWFPQP 418
           KKK IN+ R+  KK L P  +++  + L + +TF+VVMFS L+FSGFL+ LWF +P
Sbjct: 363 KKKEINRHRK--KKGLSPLFQSRAFHVLCIVVTFHVVMFSLLLFSGFLNDLWFNRP 416
```

A related GBS gene <SEQ ID 8913> and protein <SEQ ID 8914> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      3.22
GvH: Signal Score (-7.5): -4.56
     Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 7 value:  -8.55 threshold:  0.0
   INTEGRAL    Likelihood = -8.55   Transmembrane   93 - 109 (  91 - 117)
   INTEGRAL    Likelihood = -7.64   Transmembrane   21 -  37 (  19 -  39)
   INTEGRAL    Likelihood = -6.79   Transmembrane  390 - 406 ( 387 - 410)
   INTEGRAL    Likelihood = -5.20   Transmembrane   41 -  57 (  40 -  59)
   INTEGRAL    Likelihood = -2.07   Transmembrane  203 - 219 ( 200 - 221)
   INTEGRAL    Likelihood = -1.65   Transmembrane   65 -  81 (  65 -  81)
   INTEGRAL    Likelihood = -0.75   Transmembrane  125 - 141 ( 125 - 141)
   PERIPHERAL  Likelihood =  1.01     322
 modified ALOM score:   2.21

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01206(313 - 1563 of 1863)
GP|2952530|gb|AAC05775.1||AF051356(4 - 419 of 420) integral membrane protein {Streptococcus
mutans}
%Match = 50.3
%Identity = 71.0  %Similarity = 86.6
Matches = 296  Mismatches = 55  Conservative Sub.s = 65

273       303       333       363       393       423       453       483
TFDTKWEN*YQRSYERGKQVIQAFLEKLPHLDVYGNPQYFFYLILAVLPIYIGLFFKKRFALYEIIFSLSFIVMMLTGST
                       |:: ||||: |||||||||:|||||||:||||||||| |||   || |||:||||  
                    MIDFFKNLPHLEAYGNPQYFFYIILAVLPIFIGLFFKKRFPLYEAFVSLIFIVLMLTGEK
                            10        20        30        40        50        60

513       543       573       603       633       663       693       723
FNQLKSLLAYVVGQSLLVFIYKAYRKRFNHTLVFYVTVCLSIFPLFLVKLIPAISEDGHQSLFGFLGISYLTFRAVAMII
 :|: :|: |:: | : |: || |||  ::  :||: | :||:|| |||: |||  : :|||||||||||||||:| ||:
SHQIFALFFYIIWQIFCVYSYKFYRKSRDNKWIFYLHVFMSILPLSLVKITPAIWTN-QQSLFGFLGISYLTFRSVGMIM
        70        80        90       100       110       120       130

753       783       813       843       873       903       933       963
EMRDGVLKEFTLWEFLRFLLFFPTFSSGPIDRFKRFNEDYINIPDRNELLDMLGQAIHYLMLGFLYKFILAYIFGSLIMP
|||||||  ||:|||:||:|| |||||||||||:|||:||  |||::|||||| |::||:||||:|||:|| |:|::|:|
EMRDGVLTSFTFWEFIRFMLFMPTFSSGPIDRFRRFNDDYEKIPDKDELLDMLEQSVHYIMLGFFYKFVLAQILGTMILP
        150       160       170       180       190       200       210

993      1023      1053      1083      1113      1143      1173      1203
PLKELALEQGGVFNWPTLGVMYAFGFDLFFDFAGYTMFALAISNLMGIKSPINFDKPFKSRDLKEFWNRWHMSLSFWFRD
 |||:||::|| |||||||||| :|:|||||||||:|||:||||:|||||| ||::||||:|||||||||||||||||||
GLKEMALQKGGWFNWPTLGVMYVYGLDLFFDFAGYSMFAIAISNFMGIKSPTNFNQPFKSQDLKEFWNRWHMSLSFWFRD
        230       240       250       260       270       280       290

1233     1263      1293      1323      1353      1383      1413      1443
FVFMRLVKLLVKNKVFKNRNVTSSVAYIINMLLMGFWHGLTWYYIAYGLFHGIGLVINDAWVRKKKNINKERRLAKKPLL
||||||||:|||||||||||||||||||:|||:||||||:||||| ||||||:|||:||||:|||| :||||:      |
```

```
FVFMRLVKVLVKNKVFKNRNVTSSVAYIVNMLIMGFWHGVTWYYITYGLFHGVGLVLNDAWLRKKKRLNKERKAKNLSPL
          310       320       330       340       350       360       370

1473      1503      1533      1563      1593      1623      1653      1683
PENKWTYALGVFITFNVVMPSFLIFSGFLDLLWFPQPHNK**GVL*WILNQKY*QLLMTYLWRMFLL*WMKTYLTQEF*T
||| || |||: |||||||:|||||||||: |||   :|
PENGWTRALGIVITFNVVMLSFLIFSGFLNDLWFADQLSKK
          390       400       410       420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1850

A DNA sequence (GBSx1957) was identified in *S.agalactiae* <SEQ ID 5745> which encodes the amino acid sequence <SEQ ID 5746>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2611(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10143> which encodes amino acid sequence <SEQ ID 10144> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC05774 GB:AF051356 D-alanine-D-alanyl carrier protein ligase
            [Streptococcus mutans]
 Identities = 404/510 (79%), Positives = 465/510 (90%)

Query: 5   IHDMIKTIEHPAETQADPFVYDILGEVHTYGQLKVDSDSLAAHIDSLGLVEKSPVLVPGG 64
            I DMI TIE+PA+ QA+PFVY+ILGE+HTYG+LK DSDSLAAH+D L L  KSPV+VPGG
Sbjct: 6   IKDMIATIENPAQEQAEPFVYNILGEIHTYGELKADSDSLAAHLDQLDLTAKSPVVVPGG 65

Query: 65  QEYEMLATPVALTKSGHAYIPVDQHSALDRIQAIMTVAQPSLIISIGEPPLEVDNVPILD 124
            QEY MLA+PVALTKSGHAYIP+D HSALRI+AI+ VA+PSL+I++ +PP++   VP++
Sbjct: 66  QEYAMLASPVALTKSGHAYIPIDHHSALERIEAILEVAEPSLVIAVDDFPIDNLQVPVIQ 125

Query: 125 VSQVSAIFEEKTPYEVTHSVKGDDNYYIIFTSGTTGLPKGVQISHDNLLSFTNWMISDDE 184
             SQ+  IF++K  Y++ H+VKGDD YYIIFTSGTTG PKGVQISHDNLLSFTNWMI+ +
Sbjct: 126 YSQLEEIFKQKLSYQINHAVKGDDTYYIIFTSGTTGKPKGVQISHDNLLSFTNWMINAEA 185

Query: 185 PSVVERPQMLAQPPYSFDLSVMTWAPTLAMGGTLFALPKTVVNDPKKLFATINELPIQVW 244
            P+ P RPQMLAQPPYSFDLSVMTWAPTLA+GGTLFALPK +  DPK+LF TIN+LPI VW
Sbjct: 186 PATPHRPQMLAQPPYSFDLSVMTWAPTLALGGTLFALPKEITADPKQLFTTINQLPIGVW 245

Query: 245 TSTPSFADMALLSNDPNSFTLPQLVTHFYFDGEELTVKTAQKLRQRFPKARIVNAYGPTEA 304
            TSTPSF DMA+LS+DPN++ LP LTHFYFDGEELTVKTA+KLRQRFP+ARIVNAYGPTEA
Sbjct: 246 TSTPSFVDMAMLSDDFNAQQLPHLTHFYFDGEELTVKTAKKLRQRFPQARIVNAYGPTEA 305

Query: 305 TVALSAVAITDEMLETCKRLPIGYTKDDSPTYVIDEEGHKLPNGEQGEIIIAGPAVSKGY 364
            TVALSA+A+TD+MLETCKRLPIGYTK DSPT++IDE GHKL NG+QGEII++GPAVSKGY
Sbjct: 306 TVALSALAVTDKMLETCKRLPIGYTKPDSPTPIIDESGHKLANGQQGEIIVSGPAVSKGY 365

Query: 365 LNNPEKTAEAFPQFEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELEDVSQNL 424
            LNNPE+TA AFF+FEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELEE+VSQNL
Sbjct: 366 LNNPERTAAAFFEFEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELEEVSQNL 425

Query: 425 NKSQVVKSAVAVPRYNKDHKVQNLLAYIVLKEGVRDDFSRDLDLTKAIKEDLKDIMDYM 484
            NKSQY+ SAVAVPRYNKDHKVQNLLAY+VLK+GV + FSR LD+TKAIK DL+D+MMDYM
Sbjct: 426 NKSQYISASAVAVPRYNKDHKVQNLLAYVVLKDGVEEQFSRALDITKAIKADLQDVMMDYM 485
```

```
Query: 485 MPSKFIYRKDLPLTPNGKIDIKGLMSEVNK 514
           MPSKF+YR+DLPLTPNGKIDIKGLMSEVNK
Sbjct: 486 MPSKFLYRKDLPLTPNGKIDIKGLMSEVNK 515
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5747> which encodes the amino acid sequence <SEQ ID 5748>. Analysis of this protein sequence reveals the following:

```
      Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.28   Transmembrane    92 - 108 (  91 - 108)
   INTEGRAL    Likelihood = -0.85   Transmembrane    43 -  59 (  41 -  59)

----- Final Results -----
                bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC05774 GB:AF051356 D-alanine-D-alanyl carrier protein ligase
             [Streptococcus mutans]
 Identities = 365/511 (71%), Positives = 438/511 (85%)

Query: 2   IKDMIDSIEQFAQTQADPFVYDCLGERRTVGQLKRDSDSIAAFIDSLALLAKSPVLVPGA 61
           IKDMI +IE FAQ QA+FPVY+ LGE  TYG+LK DSDS+AA +D L L AKSPV+VFG
Sbjct: 6   IKDMIATIENFAQEQAEFPVYNILGEIHTYGELKADSDSLAAHLDQLDLTAKSPVVVPGG 65

Query: 62  QTYDMLATFVALTKSGHAYIPVDVHSAPERILAIIEIAKPSLIIAIEEPLTIEGISLVS 121
           Q Y MLA+FVALTKSGHAYIP+D HSA ERI AI+E+A+PSL+IA+++FP+    + ++
Sbjct: 66  QEYAMLASFVALTKSGHAYIPIDHHSALERIEAILEVAEPSLVIAVDDFPIDNLQVFVIQ 125

Query: 122 LSRIESAKLAEMPYERTHSVKGDDNYYIIFTSGTTGQPKGVQISHDNLLSFTNWMIEDAA 181
            S++E     ++ Y+  H+VKGDD YYIIPTSGTTGKPKGVQISHDNLLSFTNWMI   A
Sbjct: 126 YSQLSEIFKQRLSYQINHAVKGDDTYYIIPTSGTTGKPKGVQISHDNLLSFTNWMINAEA 185

Query: 182 FDVFKQPQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKELVADPKQLFTTIAQLPVGIW 241
           F  P +PQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKE+ ADPKQLFTTI QLP+G+W
Sbjct: 186 FATPHRPQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKEITADPKQLFTTINQLPIGVW 245

Query: 242 TSTPSFADMGAMLSDDFCQAKMPALTHFYFDGEELTVSTARKLFERFPSAKIINAYGPTEA 301
           TSTPSF DMGAMLSDDF   ++P L/HFYFDGEELTV TA+KL +RFP A+I+NAYGPTEA
Sbjct: 246 TSTPSFVDMGAMLSDDFNAQQLPHLTHFYFDGEELTVKTAKKLRQRFPQARIVNAYGPTEA 305

Query: 302 TVALSAIEITREMVDNYTRLPIGYPKDSPTYIIDEDGKELSSGEQGEIIVTGPAVSKGY 361
           TVALSA+ +T +M++     RLPIGY KPDSPT+IIDE G +L++G+QGEIIV+GPAVSKGY
Sbjct: 306 TVALSALAVTDKMLETCKRLPIGYTKPDSPTFIIDESGHKLANGQQGEIIVSGPAVSKGY 365

Query: 362 LNNPEKTAEAFFTFKGQPAYHTGDIGSLTEDNILLYGGRLDFQIKYAGYRIELEDVSQQL 421
           LNNPE+TA AFF F+G PAYHTGD+GS+T++ +LLYGGR+DFQIK+ GYRIELE+VSQ L
Sbjct: 366 LNNPERTAAAFFEFEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELEEVSQNL 425

Query: 422 NQSPMVASAVAVPRYNKEHKVQNLLAYIVVKDGVKERFDRELELTKAIKASVKDHMMSYM 481
           N+S  +ASAVAVPRYNK+HKVQNLLAY+V+KDGV+E+F+R L++TKAIKA ++D MM YM
Sbjct: 426 NKSQYIASAVAVPRYNKDHKVQNLLAYVVLKDGVEEQFERALDITKAIKADLQDVMMDYM 485

Query: 482 MPSKFLYRDSLPLTPNGKIDIKYLINEVNNR 512
           MPSKFLYR  LPLTPNGKIDIK L++EVN +
Sbjct: 486 MPSKFLYRKDLPLTPNGKIDIKGLMSEVNKK 516
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 374/510 (73%), Positives = 439/510 (85%)

Query: 4   MIHDMIKTIEHFAETQADPFVYDILGEVHTYGQLKVDSDSLAAHIDSLGLVEKSPVLVFG 63
           MI DMI +IE FA+TQADPFVYD LGE  TYGQLK DSDS+AA IDSL L+ KSPVLVFG
Sbjct: 1   MIKDMIDSIEQFAQTQADPFVYDCLGERRTYGQLKRDSDSIAAFIDSLALLAKSPVLVFG 60
```

```
Query: 64  GQEYEMLATFVALTKSGHAYIPVDQHSALDRIQAINTVAQPSLIISIGEFPLEVDNVPIL 123
           Q Y+MLATFVALTKSGHAYIPVD HSA +RI AI+ +A+PSLII+I EFPL ++ + ++
Sbjct: 61  AQTYDMLATFVALTKSGHAYIPVDVHSAPERILAIIEIAKPSLIIAIEEFPLITEGISLV 120

Query: 124 DVSQVSAIPEEKTPYEVTHSVKGDDNYYIIFTSGTTGLPKGVQISHDNLLSFTNWMISDD 183
            +S++ +    + PYE THSVKGDDNYYIIFTSGTTG PKGVQISHDNLLSFTNWMI D
Sbjct: 121 SLSEIESAKLAEMPYERTHSVKGDDNYYIIFTSGTTGQPKGVQISHDNLLSFTNWMIEDA 180

Query: 184 EFSVPERPQMLAQPPYSPDLSVMYWAPTLAMGGTLFALPKTVVNDPKKLFATINELPIQV 243
            F VP++PQMLAQPPYSPDLSVMYWAPTLA+GGTLFALPK +V DPK+LF TI +LP+ +
Sbjct: 181 AFDVPKQPQMLAQPPYSPDLSVMYWAPTLALGGTLFALPKELVADPKQLFTTIAQLPVGI 240

Query: 244 WTSTPSFALMALLSNDFNSETLPQLTHFYFDGEELTVKTAQKLRQRFPKARIVNAYGPTE 303
            WTSTPSFALMA+LS+DF    +P LTHFYFDGEELTV TA+KL +RFP A+I+NAYGPTE
Sbjct: 241 WTSTPSFALMAMLSDDFCQAKMPALTHFYFDGEELTVSTARKLFERFPSAKIINAYGPTE 300

Query: 304 ATVALSAVAITDEMLETCKRLPIGYTKDDSFTYVIDEKGHKLPNGEQGEIIIAGPAVSKG 363
            ATVALSA+ IT EM++    RLPIGY K DSPTY+IDE+G +L +GEQGEII+ GPAVSKG
Sbjct: 301 ATVALSAIEITREMVDNYTRLPIGYPKPDSPTYIIDEDGKELSSGEQGEIIVTGPAVSKG 360

Query: 364 YLNNPEKTAEAFFQFEGLRAYHTGDLGSKTDNGLLLYGGRMDFQIKHNGYRIELEDVSQN 423
            YLNNPEKTAEAFF F+G PAYHTGD+GS+T++ +LLYGGR+DFQIK+ GYRIELEDVSQ
Sbjct: 361 YLNNPEKTAEAFFTFKGQPAYHTGDIGSLTEDNILLYGGRLDFQIKVAGYRIELEDVSQQ 420

Query: 424 LNKSQVVKSAVAVPRYNKDHKVQNLLAYIVLKBGVRDDPERDLDLTKAIKEDLKDIMMDY 483
            LN+S  V SAVAVPRYNK+HKVQNLLAYIV+K+GV++ P+R+L+LTKAIK  +KD MM Y
Sbjct: 421 LNQSEMVASAVAVPRYNKEHKVQNLLAYIVVKDGVKERPDRELELTKAIKASVKDHMMSY 480

Query: 484 MMPSKFIYRELDLPLTPNGKIDIKGLMSEVN 513
            MMPSKF+YR+ LPLTPNGKIDIK L++EVN
Sbjct: 481 MMPSKFLYRDSLPLTPNGKIDIKTLINEVN 510
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1851

A DNA sequence (GBSx1958) was identified in *S.agalactiae* <SEQ ID 5749> which encodes the amino acid sequence <SEQ ID 5750>. This protein is predicted to be a histidine protein kinase (phoR). Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.64   Transmembrane    9 -  25 (   5 -  32)
   INTEGRAL    Likelihood =-11.62   Transmembrane  136 - 152 ( 132 - 164)

----- Final Results -----
                bacterial membrane --- Certainty=0.6456(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB54569 GB:AJ006392 histidine kinase [Streptococcus pneumoniae]
 Identities = 105/416 (25%), Positives = 197/416 (47%), Gaps = 56/416 (13%)

Query: 7   KKFVFLTMSILIVVVLFLFAVSNRYNQYKDEYDAYRIVKLVAKNDY---LGIPGDEPIAL 63
           + F+F+ + +  ++V+ L  + NR +    +    ++  L+A  DY   L + G   I
Sbjct: 12  RDFIFILLILLGFILVVTLLLLENRRDNIQLKQVNQKVKDLIA-GDYSKVLDMQGGSEITN 70

Query: 64  VTIDNQKMVKIQSNNTDLTNDVIEKSSLKL------LEQGKKSRKWKSFIYSIKE----- 112
           +T +   + ++      LT + +S+ S +L      +  G +   +  I  I +
Sbjct: 71  ITNNNLNDLSEV----IRLTQENLEQRSKRLNSILFYMTDGVLATNRRGQIIMINDTAKKQ 126
```

```
Query: 113 ---YKDKTYTIAIMDLASYEVPYARRFLILVPT--------IPGPCLLRAVSLYLSR--- 158
             K+     +I++L   E  Y  R LI              I G  L   V    L R
Sbjct: 127 LGLVKEDVLNRSILELLKIRENYELRDLITQSPELLLDSQDINGKYLNLRVRFALIRRES 186

Query: 159 -FIVGPVS-----TEMTRSKQ----FVSDASHELKTPIAAIRANVQVLEQ----QIPGNR 204
             FI G V      TE  +E++    FVS+ SHEL+TP+ ++++ ++ L++    +
Sbjct: 187 GFISGLVAVLHDTTEQRKEERERRLFVSNVSHELRTPLTSVKSYLELALDEGALCETVAPD 246

Query: 205 YLDHVVSETKRMEFLIEDLLNLSRLDEKRSKVNFKKLNLSVLCQEVLLTYESLAYEEEKC 264
             ++   + ET RM  ++ DLL+LSR+D    S ++ + +N +       +L  ++  + +E++
Sbjct: 247 FIKVSLDETNRMMRMVTDLLHLSRIDNATSHLDVELINFTAFTTFILNRPDKMKGQEKEK 306

Query: 265 LNDTIED----DVWIVGKESQIKQILIILLDNAIRHSLSKSAIQFSLKQARRKAILTISN 320
               + + D     +W+  +  ++ Q++  +I++NAI++S     I   +K    + IL+IS+
Sbjct: 307 KYELVRDYPINSIWMEIDTDKMTQVVDNILNNAIKYSPDGGKITVRMKTTEDQMILSISD 366

Query: 321 PGAIYSKEVMDNLFERFYQAKDDHADSLS---FGLGLSIAKAIVERHKGRIRAYQE 373
               K+ +  +F+RFY+    D A S +    GIGLSIAK I+++HKG I A  E
Sbjct: 367 HGLGIPKQDLPRIFDRFYRV--DRARSRAQGGTGLGLSIAKEIIKQHKGFIWAKSE 420
```

A related sequence was also identified in GAS <SEQ ID 9131> which encodes the amino acid sequence <SEQ ID 9132>. Analysis of this protein sequence reveals the following:

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.30   Transmembrane    9 -  25 (   4 -  33)
   INTEGRAL    Likelihood =-10.35   Transmembrane  161 - 177 ( 154 - 190)
   PERIPHERAL  Likelihood =  4.35     142

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 94/406 (23%), Positives = 190/406 (46%), Gaps = 31/406 (7%)

Query: 1   MFSDLRKKFVFLTMSILIVVLFLFAVSNRYNQVWDEYDAYRIVKLVAKNDYLGIPGDRP 60
            MF+ +R +F+ +    + +++  +  + N    Y  + +  RI+ L++ N    +PG
Sbjct: 10  MFNRIRIRFIMIASIAIFTILLSSIVGIINTARCYQSQQENRILHLISSNKGK-LPGTTE 68

Query: 61  IAL-----VTIDNQKMVKIQS-----NNTDLTNDVIEKSSLKLLE-----------QGK 98
             +     ++ D+   + S     N   L+++     S+L   E             + K
Sbjct: 69  SSKRLGTKLSDSDLSQRYYSVIFNANGHLLSSNTANISALDREAQYFARLFAKSGEEK 128

Query: 99  KSRKWWSFIYS---IKEYKDKTYTIAIMDLASYEVPYARRFLILVPTIPG-PCLLAAVSLY 155
             S + +  +YS  I +  ++   + I+D   Y          + V   FG F      +
Sbjct: 129 GSYRHQDSVYSYLLTQLPNKEKLVVILDTTFPYRSVGDLLAVSVMLAPGGFIFVVLVSL 188

Query: 156 LSRPIVGPVSTEMTRKEQPVSDASHELKTPIAAIRANVQVLEQQIPGNRYLDHVVSETKR 215
             S  ++ P       ++++P+++A HELKTP+A I AN +++E      + +      + KR
Sbjct: 189 FSGMVIKPFVQNYHKQRRFITNAGHELKTPLAIISANNELVKLMTGKSEWTKSTSDQVKR 248

Query: 216 MEFLIEDLLNLSRLDEKRSKVNFKKLNLSVLCQEVLLTYESLAYEEEKCLNDTIEDDVWI 275
             +  LI  ++ L+RL+E+   V      ++ S + Q+     ++SL   ++ K  + TI+ ++ I
Sbjct: 249 LYGLINQMTTARLESQFDVV-LHMVDFSAIAQDAAEDFKSLVLKIGKRFDLTIQPNIMI 307

Query: 276 VGKESQIKQILIILLDNAIRHSLSKSAIQFSLK---QARRKAILTISNPSAIYSKEVMDN 332
             EE  + +++ IL+DNA ++    K  ++ SL     + R++A L +SN
Sbjct: 308 KAEEKSLFELVTILVDNAMKYCDPKGLVKVSLTTIGRRRKRAKLEVSNTYLEGKSIDYSR 367

Query: 333 LFERFYQAKDDH-ADSLSFGLGLSIAKAIVERHKGRIRAYQEKDQL 377
             FERFY+  + H +    +G+GLS+A+++V+  KG I    + D +
Sbjct: 368 FFERFYREDESHNSKEKGYGIGLSMAESMVKLFKGTITVNYKNDAI 413
```

A related GBS gene <SEQ ID 8915> and protein <SEQ ID 8916> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:    17.50
GvH: Signal Score (-7.5): -2.9
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 2 value: -13.64 threshold:  0.0
   INTEGRAL    Likelihood =-13.64   Transmembrane    9 -  25 (   5 -  32)
   INTEGRAL    Likelihood =-11.62   Transmembrane  136 - 152 ( 132 - 164)
   PERIPHERAL  Likelihood =  2.49     345
 modified ALOM score:   3.23

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6456(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
28.3/57.2% over 371aa
                                                    Listeria monocytogenes
  GP|6117973| LisK Insert characterized

ORF00341(631 - 1452 of 1785)
GP|6117973|gb|AAF03933.1|AF139908_3|AF139908(105    -   476   of   483)   LisK   {Listeria
monocytogenes}
%Match = 8.4
%Identity = 28.2  %Similarity = 57.1
Matches = 79  Mismatches = 113  Conservative Sub.s = 81

459       489       519       549       579       609       639       669
VKLVAKNDYLGIPGDEPIALVTIDNQKMVKIQSNNTDLTNDVIEKSSLKLLEQGKKSRKWKSPIYSIKEYKDKTYTIAIM
  :           |  : : |           :  |          :|:  : |:    :   |   :  | |  ::  |  :
QGIGQMLLNEEEPEVKELLLATTSTLTNQDLTDNEEIKYLFNNDKTVNRKLQDQVINLYDKDGHFINKYYPSRSQDTYSI
          50        60        70        80        90       100       110

699       729                                                 756
DLASYEVFYARRFLILVFTIFG-------------------------------------FCLLAAVSLVLSRFI--
|:: | |    :|::    || |                                     || ||:|::|  :
DFSQYFVSGTDKFIMNKPTIDQQKMMTAQKPIVADDNTTVIGYAQVVNPLTSYNRMMDRLLVTMILLGVALFLSGWLGY
          130       140       150       160       170       180       190

783       813       843       873
-----------------------------------------------VGPVETEMTREKQPVSDASHELKTPIAAIRA
                                                :  :||   ::|||| ||||||:||: ':
LLAQNPLNPLTRLARTMNDIRKNGPQKRIETKTNSRDEIGELTVVFNDMMTRIETSFEQQKQPVEDASHELRTPVQIMEG
          210       220       230       240       250       260       270

918       948       978       1008      1038      1068      1098
NVQVLBQ---QIPG--NRYLDHVVSHTKRMEFLIEDLLNLSRLDEKRSKVNFKKLNLSVLCQEVLLTYESLAYEEEKCLN
 :::: :   |    :  |:  ::| :||: |::::|:|||  ::       ::  :::   ::|    :| : ||
HLKLLTRWGKDDPAVLDESLNASLTELERMKKLVQRMLDLSRAQSIQTKELQITTVNATVEQVKRNFS-VMYENFTFTL
          290       300       310       320       330       340       350

1128      1158      1188      1218      1248      1278      1308      1338
DTTEDDVWIVGEESQIKQILIILLDNAIRHSLSKSATQPSLKQARRKAILTISNPSATYSKEVMDNLFERFYQA-KDDHA
   : |:  : : ::::|||||::||:::|   : :   : : :::  : : :     |:| :| :|  |||:  |
KEDDTDLRALIQHNHLEQILIIIMDNAVKYSGDGTEVDMHVYKEQKQIHIDVRDYGKGISQEEIDKIFNRFYRVDKARSR
          370       380       390       400       410       420       430

1365      1395      1425      1452      1482      1512      1542      1572
DSLSPGLGLSLAKAIVERHKGRIRAYQEKDQ-LRLEVQLPIDGPWTNTMIN*RKNDETIFIFYW*NVIILRYFIVTNLLF
:    ||||:||| :|| : | | |  | |:  ::: ||     :
```

```
EKGGNGLGLAIAKQLVEGYIGTINAVSEVDKGTTIKITLPYIEPKSK
        450       460       470       480
```

SEQ ID 5750 (GBS34) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 9; MW 69kDa).

GBS34-GST was purified as shown in Figure 193, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1852

A DNA sequence (GBSx1959) was identified in *S.agalactiae* <SEQ ID 5753> which encodes the amino acid sequence <SEQ ID 5754>. This protein is predicted to be two-component response regulator (regX3). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1986(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04091 GB:AP001508 two-component response regulator [Bacillus halodurans]
 Identities = 98/223 (43%), Positives = 145/223 (64%), Gaps = 5/223 (2%)

Query: 2   RLLWVEDEKSIAEAIQALLADKGYSVDLAFDGDDGLEYILTYLYDLVLLDIMLPKRSGLS 61
           R+L++EDEK IA  +Q  L  +GY  D AF G DGLE      +DLVLLD+MLP+ SGL
Sbjct: 3   RILIIEDEKKIARVLQLELEHEGYETDAAFSGSDGLETFQAFAWDLVLLDVMLPELSGLE 62

Query: 62  VLKRVREAGLETPIIPLTAKSQTYDKVNGLDLGADDYITKPFEADELLARIR--LRTRQS 119
           VL+R+R      TPII LTA++    DKV+GLDLGA+DYITKPFE +ELLAR+R  LRT Q+
Sbjct: 63  VLRRIRMTDPVTPIILLTARNSIPDKVSGLDLGANDYITKPFEIEELLARVRACLRTVQT 122

Query: 120 SLIRANQLRLGNIRLNTDSHELESKESSVKLSNKEFLIMEVFMRNAKQIIPKNQLLSKVM 179
               + L    + +N  + +++     +++L+ KEF L+  F++N  Q++ + Q+++ VM
Sbjct: 123 RERVEDTLMFQELTLNEKTRDVQRGNETIELTPKEFRLLVFFINKNQQVLSREQILTNVM 182

Query: 180 GPSDMSEYNQLEVFISFLRKKLRFLKADIEIITTRGFGYSLEE 222
           G     + N ++V++ +LRKKL   +A   + T +G GY L+E
Sbjct: 183 GFDYYGDTNVIDVYVRYLRKKLSLTPA---LQTVRGVGYRLKE 222
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1853

A DNA sequence (GBSx1960) was identified in *S.agalactiae* <SEQ ID 5755> which encodes the amino acid sequence <SEQ ID 5756>. This protein is predicted to be 50S ribosomal protein L34-related protein. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5923(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC22660 GB:U32781 ribosomal protein L34 (rpL34) [Haemophilus influenzae Rd]
 Identities = 32/44 (72%), Positives = 37/44 (83%)

Query: 1  MKRTYQPSKIRRQRKHGFRHRMSTKNGRRVLASRRRKGRKVLSA 44
          MKRT+QPS ++R R HGFR RM+TKNGR+VLA RR KGRK LSA
Sbjct: 1  MKRTFQPSVLKRSKTHGFRARMATKNGRQVLARRRAKGRKSLSA 44
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5757> which encodes the amino acid sequence <SEQ ID 5758>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.5385(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 42/44 (95%), Positives = 44/44 (99%)

Query: 1  MKRTYQPSKIRRQRKHGFRHRMSTKNGRRVLASRRRKGRKVLSA 44
          +KRTYQPSKIRRQRKHGFRHRMSTKNGRRVLA+RRRKGRKVLSA
Sbjct: 1  VKRTYQPSKIRRQRKHGFRHRMSTKNGRRVLAARRRKGRKVLSA 44
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1854

A DNA sequence (GBSx1961) was identified in *S.agalactiae* <SEQ ID 5759> which encodes the amino acid sequence <SEQ ID 5760>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.79   Transmembrane  122 - 138 ( 115 - 141)
   INTEGRAL    Likelihood = -4.35   Transmembrane   19 -  35 (  15 -  40)

----- Final Results -----
               bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAP95990 GB:AE004350 conserved hypothetical protein [Vibrio cholerae]
 Identities = 79/145 (54%), Positives = 117/145 (80%)

Query: 1    MKTFVNNASKTVLSLWFGVMPTIMTVGTIALIISVSTPIFKILGTPFLPFLELLGIPRAD 60
            +++ +     +  + + FGV+P +M +GTIAL+I+  T +F +LG PF+PFLELLG+PEA
Sbjct: 314 VQSVIGSGIRNAVDMVFGVLPVVMGLGTIALVIAEYTSVFSLLGQPFIPFLELLGVPEAT 373

Query: 61   IASQTMIVGFSDMVVPSIMAAEIHSEMTRFIVATVSIVQLIYMSRTGAVILGSKIPINIL 120
             AS+T++VGF+DM +P+I+AA I +EMTRF++A +S+ QLIYMSR GA++LGS+IP+NI+
Sbjct: 374 AASKTIVVGFADMFIPAILAASIDNEMTRFVIAAMSVTQLIYMSEVGALLLGSRIPVNIV 433

Query: 121 ELFIIFIERTIISLPIIVLMAHLFF 145
           ELF+IFI RT+I+LP+I  +AHL F
Sbjct: 434 ELFVIFILRTLITLPVIAAVAHLLF 458
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1855

A DNA sequence (GBSx1962) was identified in *S.agalactiae* <SEQ ID 5761> which encodes the amino acid sequence <SEQ ID 5762>. This protein is predicted to be D,D-carboxypeptidase (dacA-2). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2443(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9485> which encodes amino acid sequence <SEQ ID 9486> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10945> which encodes amino acid sequence <SEQ ID 10946> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA67776 GB:X99400 D,D-carboxypeptidase [Streptococcus pneumoniae]
 Identities = 193/383 (50%), Positives = 282/383 (73%), Gaps = 6/383 (1%)

Query: 1   MAVDLDSGKILYEKDANKPAAIASLTKIMTVYMVYKEIDNGNLKWNTKVNISDYPYQLTR 60
           +AV+ ++GKILYEKDA +P  IAS+TK++TVY+VY+ ++NG++  +T V+ISDYPYQLT
Sbjct: 33  IAVEANTGKILYEKDATQPVEIASITKLITVYLVYEALENGSITLSTPVDISDYPYQLTT 92

Query: 61  RSDASNVPLEVKPRYTVKQLVDAAMISSANSAAIALAEHISGTESKFVDKMTAQLEKWGIH 120
            S+ASN+P+E R YTV++L++A ++SSANSAAIALAE I+G+E  FVD M A+L +WGI
Sbjct: 93  NSEASNIPMEARNYTVEELEATLVSSANSAAIALAEKIAGSEKDFVDMMRAKLLEWGIQ 152

Query: 121 DSHLVNASGLNNSMLGKNIYPKSSQNDENKMSARDIAIVAYHLVNEYPSILKTTSKSVAK 180
           D+ +VN +GLNN  LG++IYP S +++ENK+SA D+AIVA +L+ +YP +L+IT K  +
Sbjct: 153 DATVVNTTGLNNSTLGDNIYPGSKKDENKLSAYDVAIVARNLIKKYPQVLEITKKPSST 212

Query: 181 FDKDIMHSYNYMLPDMPVFRPGITGLKTGTTELAGQSFIATSTESGMRLLTVIMHADKAD 240
           F    + S NYML  MP +R G  GLKTGTT+ AG+SF+ T+ E GMR++TV+++AD  D
Sbjct: 213 FAGMTITSTNYMLEQMPAYRGGFDGLKTGTTDKAGESFVGTVEKGMRVITVVLNADHQD 272

Query: 241 KDKVARFTATNSLANYITNTYEPNLWLAKGRAYKGKEASVRDGKEQSVIAVAKNDLKVVQ 300
            + YARFTAT+SL++YI++T+   ++ +G AY+  +A V+DGKE +VIAVA  D+ +++
Sbjct: 273 NNPYARFTATSSLMDYISSTFTLRKIVQQGDAYQDSKAPVQDGKEDTVIAVAPEDIYLIE 332

Query: 301 KKNITKQNQLKINF---KKELTAPITKKKNLGKAYYVDLNKVGKGYLIKE-PSVHLVAKD 356
           +  +  Q+   + F    K + AP+     +G   Y D + +G+GY+  E PS  +VA
Sbjct: 333 R--VGNQSSQSVQFTPDSKAIPAPLEAGTVVGHLTYEDKDLIGQGYITTERPSFEMVADK 390

Query: 357 SIERSFFLKVWWNHFVRYVNEKL 379
            IE++FFLKVWWN FVR+VNEKL
Sbjct: 391 KIEKAFFLKVWWNQFVRFVNEKL 413
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5763> which encodes the amino acid sequence <SEQ ID 5764>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
            bacterial outside --- Certainty=0.3000(Affirmative) < succ>
```

```
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 176/380 (46%), Positives = 257/380 (67%), Gaps = 3/380 (0%)

Query: 1   MAVDLDSGKILYEKDANKPAAIASLTKIMTVYMVYKEIDNGNLKWNTKVNISDYPYQLTR 60
           +AVDL+SGK+LYEKDA +   +AS++K++T Y+VYKE+  G L W++ V IS+YPY+LT
Sbjct: 33  IAVDLESGKVLYEKDAKEVVPVASVSKLLTYYLVYKEVSKGKLNWDSPVTISNYPYELTT 92

Query: 61  ESDASNVPLEKRRYTVKQLVDAAMISSANSAAIALAEHISGTESKFVDKMTAQLEKWGIH 120
           SNVPL+KR+YTVK+L+ A  ++++ANS AIALAE I GTE KFVDKM  QL +WGI
Sbjct: 93  NYTISNVPLDKRKYTVKELLSALVVNNANSPAIALANSKIGGTEPKFVDKMKKQLRQWGIS 152

Query: 121 DSHLVNASGLNNSMLGNHIYPKSSQNDENKMSARDIAIVAYHLVNEYPSILKITSKSVAK 180
           D+ +VN++GL N  LG + YP +  +DEN   A D+AI+A HL+ E+P +LK++SKS
Sbjct: 153 DAKVVNSTGLITNHFLGANTYPNTEPDDENCFCATDLAIIARHLLLSFPEVLKLSSKSSTI 212

Query: 181 FDKDIMHSYNYMLPDMPVFRPGITGLKTGITELAGQSFIATSTESGMRLLTVIMHADKAD 240
           F    ++SYNYML  MP +R G+ GL  G ++ AG SF+ATS E+ MR++TV+++AD++
Sbjct: 213 FAGQTIYSYNYMLKGMPCYRBGVDGLFVGYSKKAGASFVATSVENQMRVITVVLNADQSH 272

Query: 241 KDKYARFTATNSLLNYITNTYEPNLVLAKGAAYKGKEASVRDGKEQSVIAVAENDLKVVQ 300
           +D  A F  TN LL Y+  ++  ++          K  V D  E++V  VA+N L  ++
Sbjct: 273 EDDLAIFKTINQLLQYLLINFQKVQLIENNKPV--KTLYVLDSPEKTVKLVAQNSLFPIK 330

Query: 301 KQNITKQNQLKINFKKE-LTAPITKKENLGKAYYVDLNKVGKGYLIKEPSVHLVAKDSIE 359
             +  +N + I  K  + AP++K + LG+A   D + +G+GYL  PS++L+ + +I
Sbjct: 331 PIHTKTKNTVHITKKSSTMIAPLSKGQVLGRATLQDKHLIGQGYLDTPPSINLILQKNIS 390

Query: 360 RSFFLKVWWNHFVRYVNEKL 379
           +SFFLKVWWN FVRYVN  L
Sbjct: 391 KSFFLKVWWNRFVRYVNTSL 410
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1856

A DNA sequence (GBSx1963) was identified in *S.agalactiae* <SEQ ID 5765> which encodes the amino acid sequence <SEQ ID 5766>. This protein is predicted to be penicillin binding protein 4 (pdp4) (dacA-1). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL     Likelihood =-12.58   Transmembrane  368 - 384 ( 363 - 394)

----- Final Results -----
              bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA60582 GB:X87104 penicillin binding protein 4 [Staphylococcus
      aureus]
 Identities = 117/333 (35%), Positives = 188/333 (56%), Gaps = 8/333 (2%)

Query: 5   IVSFLCILLSLTCVNSVQAEEHKDIMQITREAGY-DVKDINKPKASIVIDNKGHILWEDN 63
           I+  LC+ LS+    + A       +Q   + GY  +    +P +++ +   G +L++ N
Sbjct: 7   IIIILCLTLSLMTPYAQAANSDVTPVQAANQYGYAGLSAAYEPTSAVNVSQTQQLLYQYN 66

Query: 64  ADLERDPASMSKMFTLYLLFEDLAKGKTSLNTTVTATETDQAISKIYEISNNNIHAGVAY 123
            D + +PASM+K+ T+YL  E + KG+ SL+ TVT T  +  +S + E+SN  ++ G  +
```

```
Sbjct: 67  IDTKWNPASMTKLMTMYLTLEAVNKGQLSLDDTVTMTNKEYIMSTLPELSNTKLYPGQVW 126

Query: 124 PIRELITMTAVPSSNVATIMIANHLSQNNPDAPIKRINETAKKLGMTKTHFYNPSGAVAS 183
            I +L+ +T   SSN A +++A  +S+N  D F+  +N  AK +GM  THF NP+GA  S
Sbjct: 127 TIADLLQITVSNSSNAAALILAKKVSKNTSD-FVDLMNNKAKAIGMKNTHFVNPTGAENS 185

Query: 184 AFNGLYSPKEYDNNNATNVTTARDLSILTYHFLKKYPDILNYTKYPEVKAMVGTFYEEFFT 243
                ++P +Y +    VTTARD +IL  H +K+ P IL++T     K +  T +  T+
Sbjct: 186 RLR-TFAPTKYKDQERTVTTARDYAILDLHVIKETPKILDFT-----KQLAPTTHAVTYY 239

Query: 244 TYNVSTPGAKFGLEGVDGLKTGSSPGAAFNALVTAKRQNTRLITVVLGVGDWSDQDGEYY 303
            T+N+S  GAK  L G DGLKTGSS +A +N  +T KR   R+  V++G GD+ +  GE
Sbjct: 240 TFNFSLEGAKMSLPGTDGLKTGSSDTANYNHTITTKRGKFRINQVIMGAGDYKNLGGEKQ 299

Query: 304 RHPFVNALVEKGFKDAKNISSKTPVLKAVKPKK 336
            R+   NAL+E+ F   K +    + +  KK
Sbjct: 300 RNMGNALMERSFDQYKYVKILSKGEQRINGKK 332
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5767> which encodes the amino acid sequence <SEQ ID 5768>. Analysis of this protein sequence reveals the following:

```
        Possible site: 23

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL     Likelihood =-15.18   Transmembrane  371 - 387 ( 364 - 392)

------ Final Results -----
                bacterial membrane --- Certainty=0.7071(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA62899 GB:X91786 penicillin-binding protein 4 [Staphylococcus
          aureus]
 Identities = 119/328 (36%), Positives = 184/328 (55%), Gaps = 19/328 (5%)

Query: 6   IILTIFTPICF--SVMPLVHAEDVMDIT-----RQAGYT-VSEVNRKPKSSIVVDANSSDIL 57
            +++I   +C   S+M          D+T      Q GY  +S    P S++ V + +  +L
Sbjct: 4   LISIIIILCLTLSIMTFVAQATNSDVTPVQAANQYGYAGLSAAYEPTSAVNV-SQTGQLL 62

Query: 58  WQDNIDIPRDPASMSKMPTLYILFEELAKGKITMDTTITATPTDQAIANIYRISNNNIVA 117
            +Q NID   +PASM+K+ T+Y+  E + KG+++++D T+T T  +  + E+SN  +
Sbjct: 63  YQYNIDTKWNPASMTKLMTMYLTLEAVNKGQLSLDDTVTMTNKEYIMSTLPELSNTKLYP 122

Query: 118 GVAYPIRDLITWTAVPSSNAATVMIANYLSNNDASAFIDRVNATAKQLGMTNTHFSNASG 177
            G  + I DL+ +T   SSNAA +++A  +S N  S F+D +N  AK +GM NTHF N +G
Sbjct: 123 GQVWTIADLLQITVSNSSNAAALILAKKVSKN-TSDFVDLMNNKAKAIGMKNTHFVNPTG 181

Query: 178 AAAQAPQGYYNPTKYDLSASNITTARDLSKLLYAFLKKVPEIISFTNKSGVVHTWVGTPYE 237
            A     + PTKY       +TTARD + L     +K+ P+I+ FT +     T+  T
Sbjct: 182 AENSRLR-TFAPTKYKDQERTVTTARDYAILDLHVIKETPKILDFTKQLAPTTLAVT--- 237

Query: 238 EEFHTYNHSLPDNQFGMKGVDGLKTGSSPSAAPNAMITAKRGKTRLITIVMGVGDWSDQN 297
             ++T+N SL   +  + G DGLKTGSS +A +N  IT KRGK R+  ++MG GD+ +
Sbjct: 238 --YYTFNFSLEGAKMSLPGTDGLKTGSSDTANYNHTITTKRGKFRINQVIMGAGDYKNLG 295

Query: 298 GEFYRHPFVNALTEKGF---KDSKTLSK 322
            GE  R+    NAL E+ F    K  LSK
Sbjct: 296 GEKQRNMMGNALMERSFDQYKYVKILSK 323
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 226/382 (59%), Positives = 289/382 (75%), Gaps = 7/382 (1%)

Query: 12  LLSLTCVNSVQAEEHKDIMQITREAGYDVKDINKPKASIVID-NKGHILWSDNADLERDP 70
            + +  C + +     +D+M ITR+AGY V ++N+PK+SIV+D N   ILW+DN D+ RDP
```

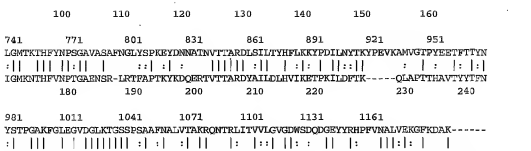
```
Sbjct: 9   IFTFICFSVMPLVHAEDVMDITRQAGYTVSEVNRPKSSIVVDANSSDILWQDNIDIPRDP 68

Query: 71  ASMSKMFTLYLLPEDLAKGKTSLNTTVTATETDQAIEKIYRISNNNIHAGVAYPIRELIT 130
           ASMSKMFTLY+LPE+LAKGK +++TT+TAT TDQAI+ IYRISNNNI AGVAYPIR+LIT
Sbjct: 69  ASMSKMFTLYILPEELAKGKITMDTTITATPTDQAIANIYRISNNNIVAGVAYPIRDLIT 128

Query: 131 MTAVPSSNVATIMIANHLSQNNPDAFIKRINETAKKLGMTKTHFYNPSGAVASAPNGLYS 190
           MTAVPSSN AT+MIAN+LS N+  AFI R+N TAK+LGMT THF N SGA A AP G Y+
Sbjct: 129 MTAVPSSNAATVMIANYLSNNDASAFIDRVNATAKQLGMTNTHFSNASGAAACAPQGYYN 188

Query: 191 PKEVDNNATNVTTARDLSILTYHFLKKYDPILNYTYPEVKAMVGTPYEETFTTVNYSTP 250
           P +VD +A+N+TTARDLS L Y FLKKYP+I+++T   V  MVGTPYEE F TTN+S P
Sbjct: 189 PTKYDLSASNITTARDLSKLLYAFLKKYPEIISFTNKSVVHTMVGTPYEESFHTYNHSLP 248

Query: 251 GAKFGLEGVDGLKTGSSPSAAPNALVTAKRQNTRLITVVLGVGDWSDQDGEYYRHPFVNA 310
             +FG++GVDGLKTGSSPSAAPNA++TAKR  TRLIT+V+GVGDWSDQ+GE+YRHPFVNA
Sbjct: 249 DNQFGMKGVDGLKTGSSPSAAFNAMITAKRGKTRLITIVMGVGDWSDQNGEFYRHPFVNA 308

Query: 311 LVEKGFKDAKNISSKT-PVLKAVKPKKEVTKTKTKSIQE--QPQTKEQWWTKTDQFIQSH 367
           L EKGFKD+K +S K    L+ + P+   TK +T S Q+  +   K+ + + + F+ +
Sbjct: 309 LTEKGFKDSKTLSKKARQKLEKLVPQ---TKKETSSKQQHFKATKKQSYLERVEDFVNHN 365

Query: 368 FVSILIVLGTIAILCLLAGIVL 389
           +LI L    I  LL  +V+
Sbjct: 366 HTFLLICLAIFIITILLLSLVV 387
```

A related GBS gene <SEQ ID 8917> and protein <SEQ ID 8918> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:    -14.02
GvH: Signal Score (-7.5): -2.54
     Possible site: 60
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value: -12.58 threshold:  0.0
   INTEGRAL    Likelihood =-12.58   Transmembrane  339 - 355 ( 334 - 365)
   PERIPHERAL  Likelihood =  1.38       99
 modified ALOM score:   3.02

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01254(301 - 1386 of 1698)
EGAD|40430|42591(32 - 419 of 431) penicillin binding protein 4 (pdp4) (Staphylococcus
aureus) GP|1125682|emb|CAA60585.1||X87105 penicillin binding protein 4 (Staphylococcus
aureus) GP|1125686|emb|CAA60582.1||X87104 penicillin binding protein 4 (Staphylococcus
aureus)
%Match = 17.3
%Identity = 36.3  %Similarity = 59.6
Matches = 123  Mismatches = 130  Conservative Sub.s = 79

264       294       324       351       381       411       441       471
FPLIFIIPDLCKLCAS*RHKDIMQITREAGY-DVKDINKPKASIVIDNKGHILWSDNADLERDPASMSKMFTLYLLPEDL
                  :|   : ||  :   :| ::: :   |::|:: | | : :||||:|: |:|| :| :
ILCLTLSIMTPYAQAANSDVTPVQAANQYGYAGLSAAYEPTSAVNVSQTGQLLYQYNIDTKWNPASMTKLMTMYLTLEAV
          20        30        40        50        60        70        80

501       531       561       591       621       651       681       711
AKGKTSLNTTVTATETDQAISKIYEISNNNIHAGVAYPIRELITMTAVPSSNVATIMIANHLSQNNPDAFIKRINETAKK
 ||: ||: ||| |  :  :| : |:|| :: |  : | :|: :|   ||| | :::|  :|:|  | |:  :|  ||
NKGQLSLDDTVTMTNKEYIMSTLPELSNTKLYFGQVWTIADLLQITVSNSSNAAALILAKKVSKNTSD-FVDLMNNKAKA
```



SEQ ID 8918 (GBS379) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 5; MW 44kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 3; MW 68.9kDa).

GBS379-GST was purified as shown in Figure 212, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1857

A DNA sequence (GBSx1964) was identified in *S.agalactiae* <SEQ ID 5769> which encodes the amino acid sequence <SEQ ID 5770>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4039(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15256 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 316/459 (68%), Positives = 386/459 (83%)

Query: 14  DLGEYKPGFHDDVKPIYSTGKGLNEAVIRELSAAKGEPEWMLDPRLKSLETFNKMPMQTW 73
           D+GEYK+GFHD     I+ + +GL + ++ E+S  K KP+WMLDPRLKSLE F  MPM  W
Sbjct: 7   DIGEYKYGFHDKDVSIPRSERGLTKEIVEEISRMKEEPQWMLDPRLKSLEHFYNMPMPQW 66

Query: 74  GADLSDIDFDDIIYYQKASDKPARDWDDVPEKIKETPERIGIPEARRAYLAGASAQYESE 133
           G DL+ ++FD+I YY K S++  R WD+VPE+IK+TP++IGIPEAE+  YLAG SAQYESE
Sbjct: 67  GGDLNSLNFDEITYYVKPSERSERSWDEVPEEIKQTPDKLGIPEAEQKYLAGVSAQYESE 126

Query: 134 VVYHNMKEEYDKLGIVFTDTDSALKSYPELFKKYFAKLVPPTDNKLAALNSAVWSGGTFI 193
           VVYHNMKE+ +  GIVF DTDSALKE  ++F+++++AK++PPTDNK AALNSAVWSGG+FI
Sbjct: 127 VVYHNMKEDLEAQGIVFKDTDSALKSNRDIFRSHWAKVIPPTDNKPAALNSAVWSGGSFI 186
```

```
Query: 194 YVPKGVKVDIPLQTYFRINNENYGQFERTLIIVDEGASVHYVEGCTAPTYSSNSLHAAIV 253
           YVPKGVKV+ PLQ YFRIN+EN GQFERTLIIVDE ASVHYVEGCTAP Y++NSLH+A+V
Sbjct: 187 YVPKGVKVETPLQAYFRINSENMGQFERTLIIVDERASVHYVEGCTAPVYTTNSLHSAVV 246

Query: 254 EIFALDGAYMRYTTIQNWSDNVYNLVTKRATAKKDATVEWIDGNLGAKTIMKYPSVYLDG 313
           EI    G Y RYTTIQNW++NVYNLVTKR    +++AT+EWIDGN+G+K TMKYP+  L G
Sbjct: 247 EIIVKKGGYCRYTTIQNWANNVYNLVTKRTVCEENAYMEWIDGNIGSKLTMKYPACILKG 306

Query: 314 EGARGTMLSIAFANKQQNQDTGAKMIHNAWTSSSIVSKSIAKGGGKVDYRGQVTFNKDS 373
           EGARG  LSIA A KQQNQD GAKMIH AP+TSS+IVSKSI+K GGKV YRG V F + +
Sbjct: 307 EGARGMTLSIALAGKQQHQDAGAKMIHLAPNTSSTIVSKSISKQGGKVTYRGIVHFGRKA 366

Query: 374 KKSVSHIECDTILMDDISKSDTIPFNEIHNSQVALEHRAKVSKISEEQLLYLMSRGLSEA 433
           + + S+IECDT++MD+ S SDTIP+NEI N  ++LEHRAKVSK+SEEQL+YLMSRG+SEA
Sbjct: 367 EGARSNIECDTLIMDNKSTSDTIPYNEILNDNISLEHRAKVSKVSEEQLFYLMSRGISEE 426

Query: 434 EATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG 472
           EATEMIVMGF+EPFTKELPMEYAVE+NRLI +EMEGS+G
Sbjct: 427 EATEMIVMGFIEPFTKELPMEYAVEMNRLIKFEMEGSIG 465
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5771> which encodes the amino acid sequence <SEQ ID 5772>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3780(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 445/472 (94%), Positives = 461/472 (97%)

Query: 1   MSEINEKVEPQPIDLGSYKPGPHDDVKPIYSTGKGLNEAVIRELSAAKGEPRWMLDPRLK 60
           MS+INEKVEP+PIDLG+Y+PGPHDDV+PIYSTGKGL+EAV+RELSAAK EPEWML+PRLK
Sbjct: 1   MSDINEKVEPKPIDLGDYQPGPHDDVEPIYSTGKGLSEAVVRELSAAKNEPEMMLEPRLK 60

Query: 61  SLETFNKMPMQTWGADLSDIDFDDIIYYQKASDKFARDWDDVPEKIKETFERIGIPEAER 120
           SLETFNKMPMQTWGADLSDI+FDDIIYYQKASDKFAR WDDVPEKIKETF+RIGIPEAER
Sbjct: 61  SLETFNKMPMQTWGADLSDINFDDIIYYQKASDKFARSWDDVPEKIKETFDRIGIPEAER 120

Query: 121 AYLAGASAQYESEVVYHNMKEEYDKLGIVFTDTDSALKEYPELPKKYPAKLVPPTDNKLA 180
           AYLAGASAQYESEVVYHNMK E++KLGI+FTDTDSALKEYP+LPK+YPAKLVPPTDNKLA
Sbjct: 121 AYLAGASAQYESEVVYHNMKGEPEKLGIIFTDTDSALKEYPDLPKQYPAKLVPPTDNKLA 180

Query: 181 ALNSAVWSGGTFIYVPKGVKVDIPLQTYFRINNENTGQFERTLIIVDEGASVHYVEGCTA 240
           ALNSA WSGGTFIYVPKGVKVDIPLQTYFRINNENTGQFERTLIIVDEGASVHYVEGCTA
Sbjct: 181 ALNSAAWSGGTFIYVPKGVKVDIPLQTYFRINNENTGQFERTLIIVDEGASVHYVEGCTA 240

Query: 241 PTYSSNSLHAAIVEIFALDGAYMRYTTIQNWSDNVYNLVTKRATAKKDATVEWIDGNLGA 300
           PTYSSNSLHAAIVEIFALDGAYMRYTTIQNWSDNVYNLVTKRA A  DATVEWIDGNLGA
Sbjct: 241 PTYSSNSLHAAIVEIFALDGAYMRYTTIQNWSDNVYNLVTKRARALTDATVEWIDGNLGA 300

Query: 301 KTIMKYPSVYLDGEGARGTMLSIAFANKGQHQDTGAKMIHNAPNTSSSIVSKSIAKGGGK 360
           KTIMKYPSVYLDG GARGTMLSIAFAN GQHQDTGAKMIHNAPNTSSSIVSKSIAK GGK
Sbjct: 301 KTIMKYPSVYLDGPGARGTMLSIAFANAGQHQDTGAKMIHNAPNTSSSIVSKSIAKSGGK 360

Query: 361 VDYRGQVTFNKDSKKSVSHIECDTILMDDISKSDTIPFNEIHNSQVALEHRAKVSKISEE 420
           VDYRGQVTFNK SKKSVSHIECDTILMDDISKSDTIPFNEIHNSQVALEHRAKVSKISEE
Sbjct: 361 VDYRGQVTFNKQSKKSVSHIECDTILMDDISKSDTIPFNEIHNSQVALEHRAKVSKISEE 420

Query: 421 QLYYLMSRGLSEASATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG 472
           QLYYLMSRGLSE+ATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG
Sbjct: 421 QLYYLMSRGLSESSATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG 472
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1858

A DNA sequence (GBSx1965) was identified in *S.agalactiae* <SEQ ID 5773> which encodes the amino acid sequence <SEQ ID 5774>. This protein is predicted to be nitrogen fixation protein (nifU). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1078(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15257 GB:Z99120 similar to NifU protein homolog [Bacillus subtilis]
 Identities = 72/139 (51%), Positives = 92/139 (65%)

Query: 4   SKLDNLYMAVVADHSKHPHHHGFLEGVEQVQLNNPTCGDVISLSVKFDGNIISDIAFAGN 63
           + LD LY  V+ DH K+P + G L     V +NNPTCGD I L++K DG+I+ D  F G
Sbjct: 5   ANLDTLYRQVIMDHYKNPRNKGVLNDSIVVDMNNPTCGDRIRLUTMKLDGDIVEDAKFEGS 64

Query: 64  GCTISTASSSMMTDAVIGKTKEEALQLADVFSKMVQGDQNPKQEKLGDAEFLAGVSKFPQ 123
           GC+IS AS+SMMT A+ GK  E AL ++ +FS M+QG +       LGD E L GVSKFP
Sbjct: 65  GCSISMASASMMTQAIKGKDISTALSMSKIFSDMMQGKEYDDSIDLGDIEALQGVSKFPA 124

Query: 124 RIKCATLSWNALRKAIERD 142
           RIKCATLSW AL K + ++
Sbjct: 125 RIKCATLSWKALEKGVAKE 143
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5775> which encodes the amino acid sequence <SEQ ID 5776>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1202(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 114/146 (78%), Positives = 133/146 (91%)

Query: 1   MALSKLDNLYMAVVADHSKHPHHHGFLEGVEQVQLNNPTCGDVISLSVKFDGNIISDIAF 60
           MALSKL++LYMAVVADHSK PHHHG L+GVE VQLNNPTCGDVISL+VKFD + I DIAF
Sbjct: 1   MALSKLNHLYMAVVADHSKRPHHHGQLDGVEAVQLNNPTCGDVISLTVKFDEDKIEDIAF 60

Query: 61  AGNGCTISTASSSMMTDAVIGKTKEEALQLADVFSKMVQGDQNPKQEKLGDAEFLAGVSK 120
           AGNGCTISTASSSMMTDAVIGK+KEEAL LAD+FS+MVQG +NP Q++LG+AE LAGV+K
Sbjct: 61  AGNGCTISTASSSMMTDAVIGKSKEEALALADIFSEMVQGQENPAQKELGEAELLAGVAK 120

Query: 121 FPQRIKCATLSWNALRKAIERDNQAE 146
           FPQRIKC+TL+WNAL++AI+R   A+
Sbjct: 121 FPQRIKCSTLAWNALKEAIKRSANAQ 146
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1859

A DNA sequence (GBSx1966) was identified in *S.agalactiae* <SEQ ID 5777> which encodes the amino acid sequence <SEQ ID 5778>. This protein is predicted to be nitrogen fixation protein (nifS) (b1680). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2453(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15258 GB:Z99120 similar to NifS protein homolog [Bacillus subtilis]
 Identities = 240/400 (60%), Positives = 306/400 (76%), Gaps = 5/400 (1%)

Query: 9   LKQDFPILNQLVNDEPLIYLDNAATTQKPNQVLEALRDYYQNDNANVHRGVHTLAERATA 68
           +++ FPIL+Q VN   L+YLD+AAT+QKP  V+E L  YY   N+NVHRGVHTL  RAT
Sbjct: 6   IREQFPILHQQVNGHDLVYLDSAATSQKPRAVIETLDKYYNQYNSNVHRGVHTLGTRATD 65

Query: 69  QYENAREKARQFLNAKLGKEILFTRGTTTGLNWVA-KFAESILERGDEVLISIMEHHSNI 127
            YE AREK R+F+NAK    EI+FT+GTTT LN VA  +A + L+ GDEV+I+ MEHH+NI
Sbjct: 66  GYEGAREKVRKFINAKSNAEIIFTKGTTTSLNMVALSYARANLKPGDEVVITVMEHHANI 125

Query: 128 IPWQQACERTGAKLVYAYLK-DGSLDLEDFYNKLSSKTKFVSLAHISNVLGCVTPVKAIA 186
           IPWQQA + TGA L Y  L+ DG++ LED     ++S TK V+++H+SNVLG V P+K +A
Sbjct: 126 IPWQQAVKATGATLKYIPLQEDGTISLEDVRETVTSNTKIVAVSHVSNVLGTVNPIKEMA 185

Query: 187 ERVHQVGAYMVVDGAQSAPHMAIDVQDLDCDFFALSGHKMLGPTGIGVLYGKESILDKMP 246
           +  H  GA +VVDGAQS PHM IDVQDLDCDFFALS HKM GPTG+GVLYGK+++L+ M
Sbjct: 186 KIAHDNGAVIVVDGAQSIPHMKIDVQDLDCDFFALSSHKMCGPTGVGVLYGKKALLEMME 245

Query: 247 PVEPGGEMIDFVKEQSATWKELPWKFEAGTPNIAGAIAFGEALDYLTDVGMDEIHQYEQS 306
           P EPGGEMIDFV    +TWKELPWKFEAGTP IAGAI  G A+D+L ++GMDEI ++E
Sbjct: 246 PAEPGGEMIDFVGLYESTWKELPWKFEAGTPIIAGAIGLGAAIDFLEEIGLDEISRHEHK 305

Query: 307 LVSYVLPKLQAIDGLTIYGPSDAESHVGVIAFNLEGLHPHDVATAMDYEGVAVRAGHHCA 366
           L +Y L + + +DG+T+YGP    E   G++ FNL+ +HPHDVAT +D EG+AVRAGHHCA
Sbjct: 306 LRAYALERFRQLDGVTVYGP---ESERAGLVTFNLDDVHPHDVATVLDAEGIAVRAGHHCA 362

Query: 367 QPLINHLGIHSAVRASFYFYNTKEDCDKLVDAIQKTKEFF 406
           QPL+  L + +  RASFY YNT+E+ DKLV+A+QKTKE+F
Sbjct: 363 QPLMKWLDVTATARASFYLYNTEEEIDKLVEALQKTKEYF 402
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5779> which encodes the amino acid sequence <SEQ ID 5780>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3714(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 293/408 (71%), Positives = 349/408 (84%)

Query: 3   LLDSYKLKQDFPIAQLVNDEPLIYLDNAATTQKPNQVLEALRDYYQNDNANVHRGVHTL 62
           LLD+  +KQDF IAQ VNDEPL+YLDNAATTQKP  VLEAL+ YYQ DNANVHRGVHTL
Sbjct: 1   LLDAKDIKQDFQIAQQVNDEPLVYLDNAATTQKPALVLEALQSYYQEDNANVHRGVHTL 60
```

```
Query: 63  AERATAQYSNARKEKARQFLNAKLSKEILFTRGTTTGLNWVAKFAESILERGDEVLISIME 122
           AERAT +YS +R++   F++AK SKE+LFTRGTTT LNWVA+FAE +L   DEVLISIME
Sbjct: 61  AERATLKYSASRQQVADFIHAKSSKEVLFTRGTTTSLNWVARFAEQVLTPEDEVLISIME 120

Query: 123 HHSNIIPWQQACERTGAKLVYAYLKDGSLDLEDFYNKLSSKTKPVSLAHISNVLGCVTPV 182
           HH+NIIPWQQAC++TGA+LVY YLKDG LD++D  NKL++KT+PVSL H+SNVLGC+ P+
Sbjct: 121 HHANIIPWQQACQKTGARLVYVYLKDGQLDMDDLANKLTTKTRPVSLVHVSNVLGCINPI 180

Query: 183 KAIAERVHQVGAYMVVDGAQSAPHMAIDVQDLDCDFFALSGHKMLGPTGIGVLYGKSSIL 242
               K IA+  H  GAY+VVDGAQS PH+AIDVQDLDCDFFA S HKMLGPTG+GVLYGK  +L
Sbjct: 181 KETAKLAHAKGAYLVVDGAQSVPHLAIDVQDLDCDFFAFSAHKMLGPTGLGVLYGKSELL 240

Query: 243 DKMPPVEPGGEMIDFVYEQSATWKELPWKFEAGTPNIAGAIAFGEALDYLTDVGMDEIHC 302
           +++ P+EPGGEMIDFVYEQ ATWKELPWKFEAGTP+IAGAI   A+ YL  +GM +IH
Sbjct: 241 NQVEPLEPGGEMIDFVYEQEATWKELPWKFEAGTPHIAGAIGLGAAISYLQRLGMADIHA 300

Query: 303 YEQSLVSYVLPKLQAIDGLTIYGPSDAESHVGVIAFNLEGLHPHDVATAMDYEGVAVRAG 362
           +E  L++YVLPKL+AI+GLTIYGPS   +  G+I+FNL+ LHPHD+ATA+DYEGVAVRAG
Sbjct: 301 HRASLIAYVLPKLRAIEGLTIYGPSQPSARSGLISFNLDDLHPHDLATALDYEGVAVRAG 360

Query: 363 HHCAQPLINHLGIHSAVRASFYFYNTKEDCDKLVDALQKTKEFFNGTL 410
           . HHCAQPL++LG+ + VRASFY YNTK DCD+LV+ALI K KEFFNGTL
Sbjct: 361 HHCAQPLLSYLGVPATVRASFYIYNTKADCDRLVEALLKAKEFFNGTL 408
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1860

A DNA sequence (GBSx1967) was identified in *S.agalactiae* <SEQ ID 5781> which encodes the amino acid sequence <SEQ ID 5782>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1441(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07189 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 171/430 (39%), Positives = 267/430 (61%), Gaps = 15/430 (3%)

Query: 1   MSKEAILNPLQAKGEPTWLQELRLKAFEKIEELELPVIERVKFHRWNLG--DGTILENDY 58
           + KE + +F  A+ EP W +++RLK FE +R LELP  ++ K   WN    D  + E
Sbjct: 9   IDKEYVQSFSDARNEPQWFKDIRLKGFELVETLELPKDKTKITSWNFTNFDHKLPSVSP 68

Query: 59  TANVPDFTE---------LGNNPKLVQIGTQTVLEQVPMELIEKGVVFTDFYSALEEIPE 109
           A++ +  +                LVQ    V  ++   L  KGV+FTD  +A++E  +
Sbjct: 69  VASIDELKDSVKGLIGEASDTQNLLVQRDATVVVYSKLDEALKAKGVIFTDLITAVKEHGD 128

Query: 110 VIERYPGK-ARPFESDRLAAYHTAYFNSGAVLYIPDNVKITQPIESLFYQDSQSKVPFNK 168
           ++E+Y+ K A   +E+RL A H A  N G  +Y+P NVEI  P++ +P+ NVEI  P++ +P+ D++    FN
Sbjct: 129 LVEKYYMKDAVKVDENRLITALHAALVNGGTFIYVPRNVEISVPLQSVFWFDTEKAGLFN- 187

Query: 169 HILLIVGKNAKVSYLERFESIGDGTERTSANISVEVIAQAGSQIKFASIDRLGENVTTPI 228
           H++++   N+ ++Y+E + S G  +E   ANI VEV A A +++ F ++D L   VTT++
Sbjct: 188 HVIIVAEDNSSITYVENYASFG--SREAVANIVVEVFAGANAKVSFGAVDNLAAGVTTYV 245

Query: 229 SRRGKHSSDATIDWALGVMNEGNVVADFDSDLIGDGSHANLKVVAASSGRQVQGIDTRVT 288
            RR    D+ ++WALG MN+GN V++  + L+GD S A+ K V+   G Q Q   T++
Sbjct: 246 VRRAHVGRDSRVEWALQMNDGNTVSENTTHLLGDNSWADTKTVSVGRGEQKQNFTTQIF 305

Query: 289 NYGCNSVGHILQHGVILRRGTLTFNGIGHIIKGAKGADAQQESRVLMLSDKARSDANPIL 348
```